THE QUARTERLY JOURNAL OF
EXPERIMENTAL PSYCHOLOGY

Vol. II 1950 Part 4

PROGRAMME FOR EXPERIMENTS ON THINKING*

BY

F. C. BARTLETT

(*From the Psychological Laboratory, Cambridge*)

In remembering, items are used in the form in which they are alleged to have been perceived at some time in the past. In thinking, however, evidence is given and the essential character of the process is to move beyond this evidence, but in line with it, to something for which so far, no exact or complete perceptual basis is claimed.

From an experimental point of view the best approach is to consider thinking as a high level skill, and to design experiments in the light of what has been discovered during recent years about the learning, practice and achievements of skill on a psycho-motor level.

When this is done, it turns out that the thinking process can usefully be considered as one in which gaps are filled up in accord with specific evidence provided. The fundamental gap-filling processes are either of interpolation or of extrapolation.

In the present paper the methodology proposed is developed and illustrated in a preliminary manner.

I

Remembering and Thinking

When I had finished the work on Remembering, and had written the book, it was in my mind that I would go on to an experimental study, perhaps somewhat similar in plan and scope, of the large variety of processes which people call "Thinking."

It had become clear that no small part of what is alleged to be recalled consists of the utilization of information which claims to have been fully presented in immediate perception at some past time or other, in the interests of the solution of a problem, or satisfaction of an interest, that belongs to the period of the recall. In fact, the information may derive from many different times and places and for some of it no immediate perceptual basis may be demonstrable. But still the claim is that what is remembered did indeed occur as it is recounted, and if items are present for which no perceptual counterpart can be found, they come in as a kind of incidental and unconsidered part of the whole process, and are not, in remembering, that which the performer is openly setting out to achieve.

Now it seemed to me, in 1930, that a tremendous and fateful further step is taken when the mind sets out specifically to use evidence having a perceptual basis in order to facilitate a step for which at the time no direct perceptual confirmation is claimed. The essence of thinking as a mental process, then, may be found in two characteristics; it uses evidence which has already accumulated through perception or recall, and it moves beyond that evidence to something which so far claims no already attained sensorial or perceptual foundation. In making this step, thinking gains many other regular characters, some of which have come to be considered widely as

* A version of a paper read on 22nd July, 1950, to a meeting of the Experimental Psychology Group in Cambridge.

ONR BRANCH LIBRARY

its defining marks. Among these the most important are, perhaps, generality and a peculiar kind of objectivity closely connected with this and best described in the Kantian manner as that which is "the same for all." But it appeared to me that probably these are characteristics which are acquired, as the essential step from evidence to beyond evidence is taken and repeated, and particularly that they are not themselves suitable characters to select as guides in the design of experiments on thinking. This was about as far as I had got at that time. It leads directly to the method of the Incomplete Story or Argument, some preliminary results of the trial of which were reported in 1938 (Bartlett, 1938) and later by D. N. Carmichael (1939; 1940).

Then came the war, and for sufficient reasons the interests of the Cambridge Laboratory were diverted predominantly to the study of varieties of psycho-physical skill, its fundamental characters and the conditions of its learning and practice. But all the time these problems of an experimental attack on the processes of thinking have been working away at the back of my mind. And they have got what is for me a new, and I hope a fruitful twist. For it now seems to me that probably the most significant thing that we can say about thinking to begin with is that it is a very high-level skill, and so before we design our experiments, we should give particular attention to what has been learned about more elementary skills in these last few years.

This has not, so far as I know, yet been done by anybody and so I have called this communication a Programme for Experiments. It is not entirely that, for I have not only planned, but to some extent tried, a few of the basic experiments. But it seems unlikely now that I shall ever be able fully to explore the field that is coming into view, and so it is one of my principal hopes that somebody else, or perhaps several others, will find the view sufficiently attractive to move close enough to it for long enough to reveal all the detail which perhaps for the present lies hidden.

II

The Character of Psycho-motor Skill

When we consider the basic character of the more elementary types of skill which have been experimentally studied we find that all of these skills consist of a sequence of moves towards some total achievement which is regarded as their completion. Any one of these moves gains its character and performs its function through its relation to the earlier and the later ones in the sequence. Even the first move of all is never an absolute beginning, but grows out of those initial circumstances which, in the psycho-motor case can be sufficiently accurately treated as static and postural relatively to what is called the beginning of the skill. We have learned that, however difficult it may be, we must sedulously avoid all treatment of the components of a skill simply as response items each complete in itself. Every component is a move from somewhere to somewhere, a step in a significant sequence. We have also learned that no mere description or measurement of the completed character of the overall achievement will give us any but speculative information about the skill. If we cannot reconstruct the steps or moves at all we may have a practical measure of what is done, but we have nothing that can throw any certain light upon learning or exercise. If we try to reconstruct the steps from the final measure only (which is what very nearly everybody, experimentalists included, *have* tried to do in the case of thinking) we can never get beyond a stage of speculation and we let ourselves in for abundance of argument. All our experiments then in the case of the mind-body skills which have been at all adequately studied have got to be laid out in a pattern

of significant sequence and where measures of the numerical kind can be used, as in these cases they can, they have to be related measures both of the final or overall achievement and of whatever our investigations show to be the dominating or important steps.

There are, as everybody will know, a number of other considerations which have turned out to be of prime importance for the experimental investigation of the psychophysiological skills, and in the past most of them have been ignored. They are set out, to my mind with almost shattering clarity, by Mr. Welford in a report on the Cambridge experiments on ageing which we hope will be published before the end of the year. There are two only of these which I must emphasize now. They are "timing," and what is—not very well, but I can't think of a better word—called "range." It seems as if the presence or absence of awkwardness in skill is principally, almost entirely, a matter of the timing of steps in their sequence. "Range" has two implications which must be carefully distinguished. It refers to measures of the limits of influence of preceding and succeeding steps or moves in the sequence, and brings into the picture in a new guise our old friend "immediate recall" and a less familiar acquaintance "immediate anticipation." It also refers to the measure of that variation of stimulus conditions which is consistent with no change in the character of a step or move relative to the overall performance.

III

Some Points in Criticism of Other Methods

These are the considerations which I now think are most highly relevant to the design of experiments on thinking, and it will be noticed that they are not derived from a theoretic analysis of the very process which it is designed to study, but from experiments upon related, but simpler processes. Before I go on to the more constructive points of this paper may I be allowed the luxury of a little criticism.

Supposing we follow the usual custom and try to choose something belonging to the final or overall achievement of thinking as our leading problem, we are almost sure to hit upon the fact that thinking is said to deal with ideas, abstraction, generalization, and so to take as our main preoccupation problems in the formation and use of concepts, as these seem to be found when the thinking mind has achieved its aim or its task. This is what nearly everybody who has attempted to discuss thinking in a psychological manner has done, whether his approach has been a philosophical phenomenalism or, as in the case of the Würtzburg people, by experiment. We shall then take or set problems, accept the solutions and try speculatively, or alternatively, by a process of reconstructive introspection, to arrive at the steps in their solutions and the conditions of these steps. We are in fact doing the very thing which lots of people tried and failed to do with any conclusiveness in the case of other kinds of skill, to understand the process in terms of a measure or assessment or analysis of its final product.

Further it is likely that we shall try to take a lot of unrelated instances and try to see how all of them illustrate the use of the same typical generalizations, abstractions, ideas or concepts. The instances will be treated, just as the composite items in a skill have often been treated, as responses each to be studied in its own right and groups of which can be arranged into classes according to their possession of common qualities, which have in fact been picked out prior to any experimentation. Few if any of the instances that are dealt with, whether by intellectual analysis or by experiment will, or can, be put into a form of significant sequence. The qualities which are

said to be found in the instances are the very qualities which are all the time taken to constitute the core of thinking, and since all of them are of a highly general nature we are extremely likely to overlook or underestimate that evidential contribution, loyalty to which has always been held to distinguish the thinker from the dealer in fancies.

There have been some exceptions to this, whether honourable or dishonourable it is hard to tell. I will briefly consider one only, because it is far the most widely known and influential. This is the conditioned reflex approach to the problems of generalization. The experiments are in the true scientific form of significant sequence, but the generalization achieved is in a peculiar position. It is found that once the conditioned response is established it remains fairly stable over a range of stimulus conditions that may cover a number of steps of discriminatable difference. Interesting as this is it is not surprising, for nowadays evidence comes from all sides to show that threshold differences do not have that direct significance for behaviour, whether bodily or mental, that was once supposed. All that may have been established may be that when an item of behaviour has been somewhat stabilized it has at once the character of varying little over a range of conditions that is not itself determined by the minimal differences which the performer is able, when he is required, to discriminate.

However this may be, the argument appears to run: Here is generalization with a small range and therefore it is the exact prototype of generalization over any range. Such arguments have been and still are very common in psychology, and I think they are never convincing and usually extremely careless. As a matter of fact when generalization occurs at the level of thinking it appears not to refer to the same or an equivalent treatment of a range of variables all in the same mode, but to an enormous and sometimes unlimited expansion to cover instances that in the way of sensory mode cannot be compared at all. In fact generalization seems to be in these cases, not a *quality* which belongs to every relatively stable response but an achievement which is reached only by some special sort of response. Indeed the character that seems most interesting in this case of conditioned response is that such generalization that is demonstrated is not greater, but very much more restricted, than that already belonging to the unconditioned items from which the whole process starts. No doubt much more could be said at this stage, but I come at once to the conclusion that we can get little or no help in the design of experiments on thinking if we begin with problems so set that the characters which we have already decided that we want to find are illustrated in their solution, but in such a manner that there are no articulated steps which anybody but the performer himself can examine to tell us how these characters come to be there. And similarly we can get no help if we begin with exceedingly simple responses that appear to be complete in themselves and study their attachment, as a whole, to conditions other than their originals. With this I come back to the Programme.

IV

Requirements for the Programme

These are the requirements:

1. We must provide evidence, and distinguish carefully between the cases and their conditions in which it is used, distorted or ignored. This means that the emphasis will lie in the beginning upon the concrete character of the process and not upon any abstract achievement. There is no other way in which the process can be shown as in a direct line of development from perceiving, recognizing and recall.

2. We must require a response which will move beyond evidence, and the move must be by a step or a series of steps which can be shown to be in a significant sequence and so that if any step changes in a given way and beyond a given range the others will also change in ways that may at least be experimentally treated as determinable.

3. When there are a number of such steps, as in the most characteristic cases there always will be, each must be set into relation in whatever manner may be possible, with earlier and later steps in the series, and with whatever is taken as the final achievement, and the final achievement must be set in relation to the steps and especially to their relative dominance.

4. To these requirements we can add what at present can hardly amount to much more than two questions: Is there anything in thinking which corresponds to the function of "timing" in the psycho-physiological skills? Is "range" in thinking the same thing as the demonstrated range of the lower or simpler types of skill?

V

Thinking as a Process of Filling up Gaps

If the fundamental requirement is taken to be that we must provide evidence set forth in such a way that a move is required beyond the evidence but in line with it, it becomes clear that what the thinker does is to close up, or partially close up, a gap of some kind. From the point of view of the material—or, to use the old word, content—the number of available gaps is infinite. But from the point of view of form there are basically two kinds only. We may have evidence on either side and a gap between, or we may have evidence and a gap succeeding or evidence and a gap preceding. The first demands interpolation and the other two extrapolation. From a methodological and a functional standpoint, in this programme all thinking is by interpolation or extrapolation, or by combinations of both.

Now it is interesting and also illuminating that the earliest known formal intelligence test was set in a pattern of interpolation. This was the Ebbinghaus Completion test and it ran straightway into difficulties because it was given a highly undefined verbal content, in which most of the steps offered the performer an unknown number of degrees of freedom. We have to look for a kind of evidence in which, granted that it is loyally accepted and used, the degree of freedom in the filling of the gap will be known, either for certain or at least in a statistical sense. We shall stand the best chance of getting on if we begin with systems whose elements are unequivocally and so to speak internally defined, and the most obvious of such systems are numerical, or spatial, or possibly temporal. We can use verbal systems if we like, but if we do we shall find that we must confine our initial dealings to words whose definitions are given and accepted and will not be changed while the gaps are being filled. This will practically confine us at the outset to proceedings of the formal logical type and may set the whole experimental scheme in an atmosphere of artificiality which we should try to avoid.

VI

Interpolation

I take the interpolation case first, because it is the easier of the two, though it is not, I think, so characteristic of what is called "thinking" normally as the case of extrapolation.

If you take as your evidence

$$\overline{a} \qquad\qquad \overline{b}$$

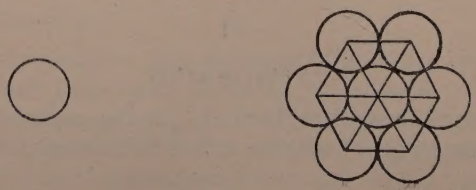
and say to your subject: There is a gap between *a* and *b*. Fill up that gap in what seems to you to be the most immediately fitting manner, so far as I have got there is 100 per cent. filling by a straight line from *a* to *b*. *a* and *b* may be near or distant, thick or thin, long or short, but, within the fairly wide limits which I have used so far, it makes no difference. Yet we cannot say simply that this is a move beyond evidence which to any normal thinking mind has one degree of freedom only. If *a* and *b* are thick they are very nearly as likely to be joined by a thin straight line as if they are themselves thin. Thickness seems in this instance to be a character of evidence from which it is rather easy to escape. There are other characters which possess the same property. We should, I think, set out to determine them systematically with this simple design of experiment.

Suppose we complicate the evidence only very slightly, beginning with something like this:—

and merely say "here is a gap: fill it," in my preliminary attempts the chances of getting a straight line joining the intersecting points *a* and *b* are about the same as those of getting a simple trellis pattern. But if we say "Here is a gap. Pay attention to the evidence that can be discovered in the two extreme figures and fill the gap accordingly," the chances of the trellis filling at once become very high indeed.

In fact one of the primary initial problems of thinking is to find out the conditions under which a normal mind will use and stay loyal to evidence provided. And I have good reason for the statement that this is a matter that can be dealt with in this design of experiment. It seems as if in a reasonably educated English population loyal use of evidence is greater in a numerical series than in a spatial series, greater in a spatial than in a temporal series, greater in a temporal than in a verbal series and vastly greater in some verbal series than in others, when the step (or steps) beyond the evidence is by interpolation.

This question of using, distorting, or neglecting evidence is not the same as that of the degrees or range of freedom open at any step, but the two are certainly related, so that the greater the freedom to step out this way, or that, or the other, the greater the tendency to reject or distort the evidence available. But it is clear that the number of different moves which must be regarded as free at any step, or any stage, of a step process of filling a gap is a question of vital importance, and it is one which this design of experiment *can* tackle. The very interesting case is where there are a number of steps in sequence and the freedom narrows from move to move until at some stage the further moves can be regarded as offering only one possibility. Suppose for example we take a case like this one:



It is obvious that the first step can be one of many. There are at least 20 possibilities, all in accord with the evidence, and there may be more. Once the first step is taken the possibilities which fit the additional evidence are drastically reduced and it seems that when the second step has been taken the remaining ones are unequivocally determined as regards their structural character, their number and their order. Objectivity in the sense of the same for all, and generality appear at a stage in a sequence.

It is then a part of this programme so to design the evidence used that at each step beyond the evidence there is a determinable amount of freedom, and that as subsequent moves are made the amount of freedom narrows down until a stage is reached of complete objectivity and universality. Only if these can be made to appear as stages in a related sequence is it likely that we can draw warrantable conclusions about their derivation.

No doubt the contrasted process may occur in which each move opens up a greater number of possible further steps. When that happens, however, we have not a thinking process but one of fancying or imagining.

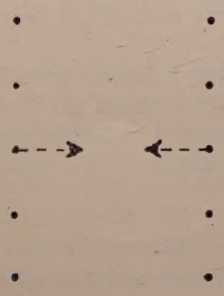
VII

Extrapolation

While as a matter of method it seems that interpolation problems are the easiest and most definite ones to illustrate, these and many other characters of thinking, nobody is likely to deny that the great bulk of what is called thinking in everyday life and in scientific investigation moves from evidence beyond evidence by some process of extrapolation. We know a result and that it is a result, but not how it has been reached. We know some of the first moves towards a result, but not the result itself. In projecting our Programme therefore we must include the extrapolation case and it is likely in fact that this will occupy a much greater place in any realization of the programme than any other device of method.

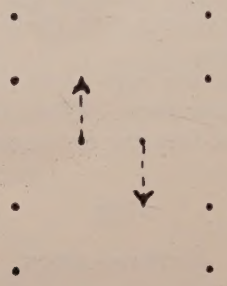
Once again, mainly for reasons of practicability, but for others as well which I will not now consider, it appears best to begin with structurally simple perceptual spatial material. We must observe our two basic requirements; we must provide evidence and the best chance that it will be used, and we must see that our experiment can proceed in moves having a significant sequence.

I will take one case suggested to me and already a good deal used by Mr. W. Adiseshiah. Here it is:



You say "Here is the first of a series of arrangements of dots. Draw the second." This is found to be quite hopeless, if, as intended, it is taken to be a problem in the extension of perceptual evidence alone. Abstractly the possibilities are infinite,

and in actual experiment they turn out to be almost unlimited. Thinking whatever else it may be, must be treated psychologically as a determinate process, and from this sort of case and a great many more, drawn from all kinds of different contents, it is very tempting to conclude in a forthright way that the mind can do nothing determinate with anything treated as only a single item of evidence. There must be at least two. Formally in the interpolation device there always are two, and it seems as if with the extrapolation method we must always provide at least two, for if we do not our performer will, and then we shall not know what the second is, except what he tells us about it. So we go on and may say "Here now is the second. Provide the third."



With this most performers will go on, with some freedom it is true, but in a determinate manner until all the dots have been moved, and whatever the order of approach, the last step is, given the character of the intermediate ones, the same for all and absolutely necessary.

It is easy to show that the number of items of evidence that have to be supplied vary with the structural character of the material, with the conventions that lie ready to be used, either of the individual or of the social group to which he belongs; and I daresay there are other conditions which will become plain as work proceeds. I have sometimes wondered what might have happened if Ebbinghaus had used extrapolation instead of interpolation as his guiding principle for intelligence testing. It could be that a more adequate measure of intelligence might be obtained in terms of the amount of evidence required, and of that kind of extrapolated gap filling which remains most meticulously loyal to the evidence provided; but as to this of course I do not know.

At this point the possibilities for the systematic development of the programme are, to me at least, as great and as fascinating as were the possibilities for pursuing my studies in recall when the early moves had been made. Who likes can explore them.

Even from the little I have done, I think I have learned, and gained at least some chance of demonstrating, something about the basic psychological character and conditions of thinking of which before I had little genuine appreciation. But it is no part of an already long communication to do more than set out the outlines of a programme.

References

1. Bartlett, F. C. (1938). The co-operation of social groups. *Occup. Psychol.*, **12**, 30–42.
2. Carmichael, D. N. (1939; 1940). The co-operation of social groups. *Brit. J. Psychol.*, **29**, 206–31; 329–44. Some examples of constructive thinking amongst Greenlanders. *Brit. J. Psychol.*, **30**, 295–315.

PERCEIVING AND NAMING A SERIES OF FIGURES

BY

K. R. L. HALL

(*From the Institute of Experimental Psychology, Oxford*)

Bartlett (1932) observed in one of his experiments that his subjects, when presented with a series of figures, tended to name them with reference to their general trend of development, sometimes anticipating the outcome rather than making an immediate identification. Two recent experiments have studied the effect of verbal "set" upon perception, and the present study was designed to show both the effect of the objective frame of reference and of verbal suggestions upon the naming of a series of figures. A series of 10 figures was presented to one group with instructions to name them. A second group was shown the same figures in exactly the reverse order. The distribution of responses was significantly different for the two groups, showing the effect of the different settings. Other groups were given the same figures, but with preliminary suggestions as to what they were to see, and their responses show some modifications. Individual variations in quality of response are also discussed, and the general implications of the method and materials for research on conceptual thinking are briefly surveyed.

I

Introduction

Bartlett (1932) had found, in one of his experiments on Perceiving, that where a series of figures was shown to his subjects which showed some particular trend of development, some tendency to "build up" to a complete figure, they were inclined to anticipate the nature of the change even in their responses to a comparatively schematic outline. After two figures of one of his series had been shown, the subjects said: "It will be a crown," thus inferring the outcome rather than responding to the immediate presentation. Bartlett described this (p. 32) as the setting up of a specific attitude, or expectation by which the perceptual act was directed and determined. He quoted this as a good example of the way the scheme or setting of the material can be used, not so much as an aid to immediate identification, but as a basis for inference such as is similarly found in other experiments that he describes.

In Bartlett's experiment, the primary task given to his subjects was that of reproducing the drawings, and the verbal reactions were subsidiary, so that the present experiment was designed along the same lines but with the purpose of studying specifically the verbal reactions to a series, and the effect on the reactions of altering the scheme or setting. It was felt also that, by making the theme of the series more complex, the variety of response might be of particular interest, in showing how individuals differed in their conceptual facility or rigidity.

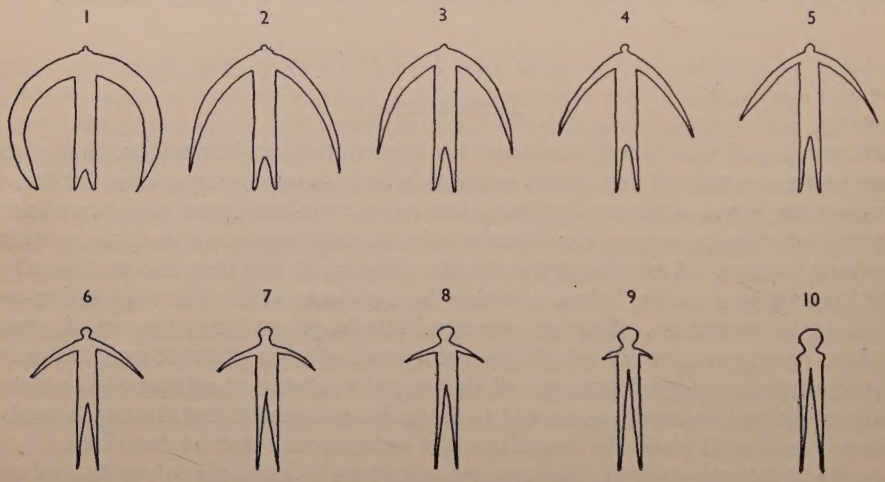
Some more recent experiments, notably those of Zangwill (1937) and Luchins (1945), have been concerned with the effect of a verbal "set" or "suggestion" upon the recognition of figures. Zangwill found that change of verbal "set" made the correct recognition of a figure considerably more difficult, while Luchins, with schoolchildren as subjects, found that the influence of a verbal suggestion depended upon what he called the structural clarity of the figure shown. In other words, when a figure was ambiguous, a suggestion was likely to be effective, and when a figure was clearly representational, then any contrary suggestion would be ineffective.

The present series of figures was devised so that only twc, possibly three, would be likely to be unambiguous, and the remainder were not intended as clear representations of any particular conventional object. Whereas Zangwill had been concerned with the way in which a name might make the identification of a figure harder, the purpose of the second part of the present experiment was to find out whether a preliminary verbal suggestion would affect the general trend of the naming of the series by significantly increasing the number of responses falling into the suggestion category.

II

Method

A series of figures was devised so that the "familiar" or unambiguous figures were in the middle and at the end of the series. Beginning in Figure 1, with a considerable distortion, the figure gradually becomes recognizable as that of a man at Figure 5, and then is again distorted away from the man-representation and becomes like a clothes' peg or some mechanical object, such as a pair of pincers. This progression is clear from the actual drawings shown below, and the distortion was made as nearly as possible equal for each stage of the series, the "arms" being reduced by ½ cm. in Figure 2, and the "legs" increased by the same amount, and so on for each successive figure of the series.



Figures 1–10

The general task for the subjects was to name each figure as it was shown to them and there was no fixed time limit for any given exposure. The subjects who carried out the experiment were all adults, mainly university students, and of the total of 140, 28 were women.

The responses given were recorded verbatim, each card being presented singly at about arm's length from the subject, who was not allowed to turn the card.

In Part I of the experiment, the problem was to discover the effect of the serial setting upon the naming of the figures, and the method used was to present the 10 figures of the series to three groups of subjects, the setting being changed for each group. The 10 subjects of Group I were shown the figures in the order shown, that is beginning with No. 1 and continuing down to No. 10. The 10 subjects of Group II were shown the figures in exactly the reverse order, beginning with No. 10. The third Group, of 80 subjects, were only shown *one* of the figures, 10 of them giving

responses to Figure 2, 10 to Figure 3, and so on. In this way, it was intended to demonstrate the effect of individual, as compared with serial, presentation of the figures upon the responses made.

In Part II, it was intended to show how a verbal suggestion would affect the distribution of responses, so the same 10 figures were used, this time with a specific verbal "set" contained in the instructions. Four more subject Groups were used. Two of the Groups were told that the first drawing they were to be shown was that of a Man, and one of these Groups was then given the figures in the order 1–10, the other in the reverse order 10–1. Two further Groups were told that the first drawing they were to be shown was that of a Penguin, and the order of presentation was 1–10 for one Group, 10–1 for the other. In this way it was hoped that any suggestion effect could be clearly demonstrated.

One difficulty at first, in view of the rather surprising diversity of responses given to the figures, was to devise a simple method of scoring and classifying the content of the responses. As the experiment proceeded, however, certain categories were used in order that the main points of difference between the various subject-groups could be conveniently tabulated. Most of the responses, as was to be expected, had some reference to a man, or some sort of human adult figure, and all such responses were classed as H. The clearest examples of such responses were given to Figures 5, 6 and 7, for instance: "A man—arms are shorter," "This one is a better human form," "A skier about to jump."

A second category of responses was headed H–. Here, there is still some human reference, but qualified in certain ways, as for example, for Figures 7, 8 and 9, a common response was: "Like a child," "Like a doll or puppet," while at the other end of the series, the qualification is more with reference to the ape or bird-like aspects of the figure, as, for instance, to Figure 3: "An ape-man," "A bird-man."

In addition, the more ambiguous figures such as 1, 2, 8 and 9, were named in a great variety of ways and classification of these responses did not seem relevant to the main purpose of the experiment. For instance, names given to Figures 1–4 included "Penguin," "Anchor," "Palm-tree," "Ornamental Arch," and so on, while to Figures 8, 9 and 10, the reference was frequently to some implement, such as "Clothes' peg" or "Pliers." A few subjects contented themselves with saying "Nothing at all," or "Nothing, though like something," these negative responses occurring particularly where the figures were presented in the order 10–1 and being probably due to the fact that once the central Man figures were passed, there is no longer any clear trend of development in the series. Many other qualitative differences occurred in the responses which are not classifiable, and some of the more striking of these will be examined later.

III

Results

As a preliminary, it was thought advisable to find out whether there was any sex difference in the type of response given to the figures, as, if there were, this would obviously affect the composition of the subject-groups. The first group of subjects, therefore, consisted of 5 male and 5 female subjects, and it was found that there was no statistically significant difference between the two sets of responses. This was confirmed also by observation of the responses in other groups tested.

We shall now first examine the results of Part I of the experiment, as to the effect of the serial setting upon the responses.

Part I

In comparing the responses given to the figures under the three different settings, order 1–10, reverse order 10–1, and individual presentation, certain clear differences in the distribution of the responses are found. The greatest difference is that between the responses to the figures when the order of presentation is reversed, while the responses given to the figures when presented singly and out of the serial context tend to lie intermediate to those of the two serial distributions. In assessing the effect of the three settings, as also in Part II of the experiment, only the responses classifiable as H or H– are taken into consideration, differences in the other responses being negligible. For this reason, in Table I below, and in the following tables, only the H and H– frequencies are shown, the balance to make up the total of 10 responses for each figure consisting of other names such as "Bird" or "Anchor," having no human reference. The table shows the differences both in the totals and in the shape of the distributions that appear to be due to the effect of the various settings, the frequency of the H and H– responses being shown for each of 10 figures.

TABLE I

Subject Group	*Serial Order*	*Response Category*	*Figure Number*										*Total*
			1	2	3	4	5	6	7	8	9	10	
I	1–10	H		2	7	9	10	8	7	3	1		47
		H–			1				2	4	1		8
		Total		2	8	9	10	8	9	7	2		55
Non-serial group (figures presented singly)		H			3	4	7	10	7	6	3		40
		H–				3	2		1	2	5		13
		Total			3	7	9	10	8	8	8		53
II	10–1	H	1			2	4	6	8	6	6		33
		H–			3	3	2	1	2	4	4		19
		Total	1		3	5	6	7	10	10	10		52

In estimating the significance of the differences shown above, it was found that, in comparing the Group I and Group II frequencies, the total number of H responses for the series was significantly greater in Group I, the value of chi square being 4·080. The probability of this occurring by chance is less than 0·05. The combined totals, however, of both the H and H– categories are not significantly different, chi square being 0·181, so that the difference in presentation order appears to affect only the relative proportion of H and H– responses, increasing, when the order is 10–1, the number of responses such as "Child" or "Doll," and decreasing those having an adult human reference.

So also, from comparing the actual distributions of the H and H– responses in the series, for each figure, the value of chi square is 14·770, which, for a 2×5 table, gives a probability of less than 0·01, while the (H plus H–) distributions do not differ significantly, the value of P being greater than 0·05.

From inspection of Table I, it appears that these differences in the type of response found in the two Groups are most marked towards the extremes of the distributions,

being smallest at Figure 7, and greatest at Figures 3 and 9. The difference in presentation order seems to have affected the responses by increasing the frequency of the H responses given to the first few figures shown.

It seems probable, therefore, that this tendency is due largely to the anticipatory nature of the responses given to the figures leading up to the comparatively familiar Figures 5, 6 and 7. In this connection, it is also of interest to note that there is a definite tendency for responses given to Figure 9 by the Group I subjects to anticipate the further development of the figure by calling it a "Clothes' peg" or "Pliers," whereas all 10 of the responses given to the same figure when shown as second of the series (i.e. to the Group II subjects) were either H or H–.

With regard to the responses given to the figures when presented singly to subjects, neither the totals for H nor the distributions of H along the series are significantly different, as found from comparing these Non-serial responses both with Group I and Group II. As was observed from Table I, the non-serial responses tend to lie midway in frequency between the two serial Groups.

From these results, therefore, obtained from the two serial Groups where the figures are shown in reverse order, and from the Non-serial responses obtained, there appears to be definite evidence that the names given to some of the figures depend directly upon the setting provided by the order of presentation. The same figure, when shown as No. 4 in the series, may be called "Man" 9 times, when shown as No. 7, only twice, and when shown singly, four times out of ten. So also to a greater or lesser degree with the other figures, depending on the degree of their ambiguity or familiarity.

Part II

In general, the effect of preliminary verbal suggestions upon the distribution of the responses to the figures was slight, and these instructional "sets" appear, on the whole, to have made certain local modifications of response to a few of the figures.

TABLE II

Group	*Category*	*Figure Number* 1	2	3	4	5	6	7	8	9	10	*Total*
III "Man" suggestion	H	1	2	6	9	9	9	5	3	1		45
	H–		1	1			1	4	5	4		16
	Total	1	3	7	9	9	10	9	8	5		61
I No suggestion	H		2	7	9	10	8	7	3	1		47
	H–			1				2	4	1		8
	Total		2	8	9	10	8	9	7	2		55
V "Penguin" suggestion	H		1	7	9	10	9	7	1			44
	H–		1	2			1	3	7	6	1	21
	Total		2	9	9	10	10	10	8	6	1	65

This effect can be seen, first, from the figure by figure distributions of the H and H– responses for the two suggestion Groups III and V, where the suggestions were made relative to Figure 1, and the series presented in the order 1–10. From inspection of

Table II, it will be seen that there appear to be a greater number of H– responses given to Figures 6–9 by the suggestion Group subjects than by those of Group I. This tendency is particularly marked in Group V.

Testing the significance of these observed differences, the value of chi square found from comparing the frequency of H– responses for Figures 6–9 given by Group III as against those of Group I was 3·950, with a P value of less than 0·05. In comparing the same frequencies given by Group V for the last five figures with those for Group I, the difference is greater, chi square being 6·861, with a P value of less than 0·01. Comparison of the frequencies for the two suggestion Groups themselves reveals no significant difference, so that it seems as though the Suggestion effect were non-specific, and serves to increase the number of H– responses to the figures at the end of the series. In other words, a significantly greater number of responses referring to inanimate objects, such as "Clothes' peg," "Pincers," etc., is given particularly to Figures 8 and 9 when there is no verbal suggestion.

The suggestions, therefore, so far from increasing the number of responses of the particular category to which they refer, seem to have had the general effect of reducing the anticipation found in the Group I responses where the subjects tended to look forward to the outcome by calling, for instance, Figure 8 a "Clothes' peg." This is, perhaps, rather a surprising result in view of the fact that the suggestions were only made relative to the first figure to be shown, and it indicates that it is the attitude of the subjects to the task as a whole which is altered, for, once the "theme" of the series is reached with the "Man" figures at Nos. 5 and 6, the possibility of a further development to either another "Man" figure, or to a "Penguin" figure, is very remote.

The main interest of this result is, therefore, in showing how a very limited suggestion may continue to be effective throughout the task, how it alters the setting and acts as an additional frame of reference to the subjects' responses. It will further be of interest to note whether any comparable effect is found with the other two suggestion Groups, IV and VI, where the figures were presented in the reverse order 10–1.

The distributions are shown in Table III below:

TABLE III

Group	*Category*	*Figure Number*										*Total*
		1	2	3	4	5	6	7	8	9	10	
IV "Man" suggestion	H				2	4	5	9	3	3		26
	H–			2	1			1	6	4	1	15
	Total			2	3	4	5	10	9	7	1	41
II No suggestion	H	1			2	4	6	8	6	6		33
	H–			3	3	2	1	2	4	4		19
	Total	1		3	5	6	7	10	10	10		52
VI "Penguin" suggestion	H		2			1	4	6	5	3		21
	H–	1			1	1	2		3	4	1	13
	Total	1	2		1	2	6	6	8	7	1	34

In comparing the distribution of the H and H– responses for Groups IV and VI with the No-suggestion Group II, we again find that the differences are not specific, that is, there is no increase in the number of H and H– responses in the "Man" suggestion, nor is there any significant increase in the number of "Bird" or "Penguin" responses given by Group VI.

With regard to both Suggestion Groups, the most obvious difference is the reduction in the total number of H and H– responses, while the actual shape of the distributions remains very much the same, with a tendency of the H responses to bunch towards Figure 10.

Testing the significance of the differences between Group IV and Group II, it was found that, while the difference between the totals of the H and H– scores is not significant, Group IV give significantly fewer H– responses on Figures 3–7, the number of H responses being the same (chi square is 4·355, giving a P value of less than 0·05). Further, for Figures 8 and 9, Group II give a greater number of H responses than Group IV. The effect of the "Man" suggestion therefore seems to be that of reducing the number of H responses to Figures 8 and 9, while, after the maximum is reached at Figure 7, it has had the effect of reducing the number of H–. A further point is that for Figures 1–3, Group IV give eight responses with no content at all—i.e. "Nothing," as compared with two for Group II.

There seems little doubt that the suggestion has changed the expectancy, the "setting" of the figures, so that the general scheme of development is modified in a very similar way to that observed with Suggestion Groups III and V.

In Group VI, the frequency of H and H– responses is also much reduced, comparison of the totals (H plus H–) for the series with the total for Group II being significantly different, the value of chi square being 6·608 (P less than 0·05). The distribution of the responses is not, however, significantly different. This reduction of H and H– responses was not brought about by any significant increase in the "Penguin" or "Bird" responses, but, as with Group IV, the suggestion effect seems to act unspecifically in modifying the general scheme or setting of the figures, in arousing an initial expectancy which acts as a supplementary background to the series as they are presented, and in relation to which each one is judged.

The general conclusion from both Groups III and V, IV and VI, is clear. In no case does the suggestion serve to make any significant positive increase in the number of responses of the particular category, whether "Man" or "Penguin," to which it refers. In every case there is some suggestion effect, for it modifies the setting or frame of reference of the figures as they are serially presented. The expectation, the nature of the task, is altered, anticipation of what is to come is changed.

Individual and qualitative differences.

In the responses to the various figures of the series, and in individual records, there is a considerable variety of qualitative difference. Certain of the qualities observed are very similar to some that are classified in Rorschach records, and assessed in terms of fluency, flexibility, good and poor originality, ability to see movement in the forms, and so on. While it was not the main purpose of this experiment to estimate the quality of the subjects' responses with reference to any objective criterion, it seems of interest to note some of the more striking differences in "verbal ability" shown in relation to such simple perceptual material. We shall very briefly, therefore, give some examples from the records.

(1) *Responses in which definite movement or mood is expressed.*

Particularly for the three Groups II, IV and VI, to whom the figures were shown in the order 10–1, a number of responses were given in which some activity or mood was described. For instance, Figure 2 was described as: "Someone—a man—absolutely tired out," Figure 5 was thought by one subject to be "Rather like a man on a ski jump," and then "more like a shop assistant wangling something under the counter." Figure 6 also produced some good "Movement" responses, such as: "A man picking his way through a muddy path, walking on stepping stones," and "A Gym demonstrator—definitely alive." Figure 7 in particular seemed to suggest many varieties of human movement, such as: "A man leaning over the top of a swimming bath, about to dive," "More like a man, holding his arms as if someone was about to help him on with his coat," and "A man floating on his back in the sea."

Figures 8 and 9 suggested "Someone doing a ski-jump," "A child learning to walk," "A man on stilts in a circus," "Caricature of a baby taking its first steps—rather unsteady on its pins," and "A child walking with his legs bandaged up."

There is, in these responses, a certain vividness and appropriateness which distinguishes them from the more common type of response such as "More like a man this time," or "A human figure" and so on. There seems little doubt, further, that this quality of response, so similar to the human movement response category on the Rorschach, is typical of a certain higher intellectual ability as judged, not from any I.Q. criterion, but from the academic records of some of the individuals concerned.

(2) *Further instances of perceptual verbal discrimination.*

In addition to the "movement" and "mood" qualities described, there is an originality of description about some responses which makes them seem peculiarly fitting. Some subjects appeared always content with merely nominating the figure, "Man", "Gorilla," "Bird," whatever it might be, while others sought for a formula which would describe the peculiarities and individuality of the particular figure. For example, one subject of Group IV described Figures 5–1 as follows: "More like a pick," "King Penguin," "King-Kong," "Nothing," "Nothing." Another subject from this Group described the same figures as: "Like a shirt hanging up on a line," "Like an all-in wrestler or gorilla," "A pin-headed chimpanzee," "A bird," "That's an instrument carpenters use—a bowsaw."

Perhaps in contrasting two full sets of responses, this conceptual difference will come out more clearly. In one of the sets it will be seen how fixed and unable to shift from one concept to another the one subject is, and how fluently and easily the conceptual transitions take place in the other record.

Figure	*Subject 1*	*Subject 2*
10	No, it's a clothing peg.	More like a clothes peg.
9	A child walking with his legs bandaged up.	Silhouette of someone on stilts.
8	A child learning to walk.	Like something with wings—a moth.
7	A man leaning over the top of a swimming bath, and about to dive.	More like a bird—gull's wings.
6	A man picking his way through a muddy path, walking on stepping stones.	Bird—plan view—in flight.
5	That reminds me of a ski-lift hook.	Same bird, more from the rear.
4	That reminds me of an ordinary naval anchor.	Same bird—but more from the rear.
3	Sort of like a crossroads, only not one—a roundabout.	Same bird—even further back.
2	Someone—a man—absolutely tired out.	Same bird—even further back.
1	Something to do with a conjuror's sort of stooge.	Same bird.

No doubt the verbal suggestion "Penguin" contributed to the stereotypy of record 2, and that, in itself, points to the importance of individual variations even in the perceiving and naming of such schematic figures as these. Some subjects accept both the verbal formula and the pressure of the preceding stimulus or stimuli of the series very passively, others are active in their discrimination and in their anticipation.

IV

Discussion

The results from the first part of this experiment seem to demonstrate very clearly the effect of the serial setting upon the perceiving and naming of the figures. The verbalizations of the subjects show the relational character of their reactions, and the dependence of their judgment of an individual figure upon the preceding figures of the series. There is, as Bartlett observed, an element of anticipation in the responses, but it is much more implicit than he appears to suggest, the scheme of development forming a total setting in relation to which the subject makes his responses. As with the lifting of a series of weights, where the subject has to judge whether a particular weight is lighter or heavier than the one preceding, so with the present figures, the subject builds up a standard against which each new figure is compared. A figure is "more like" a man than the preceding one, or it ceases to be a man altogether and becomes something quite different.

The general trend, or the "scheme" as Bartlett calls it, has much the same effect as the "objective set" of the Gestalt Psychologists which Cole (1939) describes as a ". . . dynamic factor, subjectively experienced as 'expectancy' (which) will influence the present organization. Thus the exposure of a series lines of of increasing lengths will produce an overestimation of the final one if the objective length of the latter is slightly less than the series 'implied.' Whether any such trend is experienced or not, the just-preceding activities—as they fade—provide a background upon which the present stimulus must fall" (p. 494).

The same interpretation must be made for the response modifications observed in the Suggestion Groups. The preliminary verbal "set" serves to arouse an expectancy with regard to the first figure to be shown. Yet although this does not appear on the first response, it continues to act as a background for the judgment of later figures in the series, so that it must modify the attitude of the subject to his perceiving and naming of the more ambiguous and uncertain figures. It must modify the effect of the trend. The verbal suggestions appear to act as a kind of latent "anchorage-point," as Koffka (1922) graphically expresses it.

The importance of the present observations seems to be not merely that they give an experimental confirmation and elaboration of what Bartlett himself had observed. Whether one chooses to speak of the underlying influence here as the "scheme," the "objective set," or the "idée directrice" (to go back to what appears to be the first work of this kind by Binet in 1900 on the general problem of "Suggestibility"), there seems little doubt as to its fundamental significance. Particularly, however, because of the nature of the study of which this present experiment formed a part, it is the possible bearing of these results upon the problem of conceptual thinking that may be of most interest.

We have seen that individuals, even of a relatively homogeneous group of subjects respond very variedly to this series of figures. In some, the perceiving and judging of the figures is lively, very quickly anticipating the development, as it were going

ahead of the objective trend and very easily unanchoring themselves from the previous background. In others, the naming is inert, the present judgment being tied to the previous background, whether that of the preceding figures or that of the verbal suggestion. The concepts alter with difficulty. So, it seems, this method and this type of graded material may well be useful in assessing at least one aspect of the conceptual impairment in organic and schizophrenic patients, who frequently are found to have great difficulty in shifting from one frame of reference to another. The nature of the schizophrenic deterioration, as compared with the organic, may become clearer through experimental study with this and other types of material, verbal and non-verbal, devised with reference to the basic findings of Bartlett and the Gestalt Psychologists.

I would like to thank Mr. R. C. Oldfield of the Institute of Experimental Psychology, Oxford, for his help and advice in carrying out this study.

References

1. Bartlett, F. C. (1932). *Remembering.* Cambridge.
2. Binet, A. (1900). *La Suggestibilité.* Paris.
3. Cole, L. E. (1939). *General Psychology.* New York.
4. Koffka, K. (1922). Perception. *Psychol Bull.*, **19**, 531–585.
5. Luchins, A. S. (1945). Social influences on perception of complex drawings. *J. Soc. Psychol.*, **21**, 257–273.
6. Zangwill, O. L. (1937). A study on the significance of attitude in recognition. *Brit. J. Psychol.*, **28**, 12–17.

(*Manuscript received 6th March,* 1950.)

THE INFLUENCE OF INTERPOLATED RECALL UPON RECOGNITION

BY

EUNICE BELBIN

(*From the Psychological Laboratory, Cambridge*)

An experiment was designed to show how immediate recall may affect recognition. A number of subjects were shown a picture. Some were asked to recall it and were then given a recognition test. Others were given only the recognition test after the same interval. Only 4 of the 16 subjects who had recalled it identified it; whereas 14 of the 16 others did so. A second experiment gave similar results.

Recall was constructed round dominant items of the picture. This distribution of emphasis together with the acceptance of an invented detail as genuine were the common causes of errors in subsequent recognition. Both the dominant and invented items in recall were those which became most obviously merged into an organization of related experiences and in consequence those which militated against subsequent recognition.

The recognition test was applied in two further groups of 16 subjects with a change in one of the dominant details (i.e. the wording). The number of subjects who now correctly identified the remainder of the material was 9 when there was no intermediate recall, and nil when immediate recall of the original material was interposed.

I

Introduction

This experiment was designed to investigate the influence of an interpolated activity upon the recognition of an original experience, when the interpolated activity was the recall of that experience.

In previous attempts to study the interrelationship of remembering and recognition of an experience, subjects have usually been aware of the experimental nature of the experience. Secondly, much of the test material (e.g. nonsense syllables and forms) has not been conducive to a study of those theories of remembering or recognition which postulate the retention of active organizations of experiences and the subsequent reorganization resultant on new and related experiences. Postman, Jenkins and Postman (1948) suggest that the non-facilitating effect of recall on subsequent recognition is due to the fact that weak associations may drop out during a test of recall and the level of subsequent recognition is thus lowered. In order to investigate this explanation with material more nearly related to every-day life than newly-learned nonsense syllables, the experiment to be reported made use of pictures which had been perceived incidentally.

II

First Experiment

Method.

Subjects were left sitting in a waiting room for 2 minutes and were given no instructions. At the end of this time, the Experimental Group were asked to recall as much as they could of a poster (X) which had been on a wall facing them 12 feet away. They were prompted by standardized questions.

The Control Group were occupied with an unrelated task during an equivalent amount of time.

A recognition test was then presented to both Groups, the experimental plan being as follows:—

Experimental Group.

Group A. Display of Poster X. Recall of Poster X. Recognition Test (Poster X).

Group B. Display of Poster X. Recall of Poster X. Recognition Test (Poster Y).

Control Group.

Group C. Display of Poster X. Unrelated Task. Recognition Test (Poster X).

Group D. Display of Poster X. Unrelated Task. Recognition Test (Poster Y).

The pictorial details of Posters X and Y were identical. They depicted a boy running across a road between two cars. The wording was changed, however, from

"Pedestrians—he may have faulty brakes" on Poster X to

"Motorists—It's not always your fault but . . ." on Poster Y.

64 Subjects were used, each of the 4 groups consisting of

5 Naval Ratings.
4 fourteen-year-old children.
7 University students.

16

In the recognition tests each subject was merely asked to state whether the picture (irrespective of the wording) was identical with the one he had originally seen and to qualify his Yes/No answer on a 3-point scale of certainty.

Results.

It will be seen from Table I that whereas 14 subjects of the Control Group correctly identified the original, only 4 of the Experimental Group were able to do so. Of the other 12 in this latter group who stated that it was not the same picture, 8 were Very Sure or Sure of their decision and 4 were Doubtful.

TABLE I

	Number of responses	
	Correct	*Incorrect*
Experimental Group A (Recognition test preceded by recall) ..	4	12
Control Group C (Recognition test only)	14	2

A correct response is scored when the subject says "Yes, it is the same picture."

There is thus a significant difference between the frequencies of subjects whose responses were correct and incorrect in these 2 groups, P being ·0008, when the method of Whitfield (1947) was used.

Similar differences are to be found by comparing the results of Experimental Group B with Control Group D (Table II) where a correct response is scored when the subject says "Yes, it is the same picture, irrespective of the wording." (All these subjects agreed that the wording was different.)

TABLE II

	Number of responses	
	Correct	*Incorrect*
Experimental Group B (Recall of Poster X followed by presentation of Poster Y)	0	16
Control Group D (Identification only of Poster Y)	9	7

There is again a significant difference between the frequencies of subjects' scores in the Control Group and the Experimental Group, due presumably to the effect of recall of the picture between the processes of perception and recognition. [P = ·001.]

III

Second Experiment

TABLE III

	Number of responses			
	Correct		*Incorrect*	
	Experimental Group	*Control Group*	*Experimental Group*	*Control Group*
Time interval between presentation of stimulus and recognition test:—				
10 minutes	1	5	4	–
½ hour	1	2	2	1
3 hours or over	–	3	4	1
	2	10	10	2

An experiment was conducted on a further group of 24 subjects using a newspaper photograph as the display with a time interval between recall and identification. A photograph of a visit to a London hospital of an eminent person was presented to all subjects who were asked casually if they had seen it in the newspapers. The Experimental Group E were asked to recall the picture in detail. At varying time intervals they were given the same picture and were asked to state whether or not this was the one they had originally seen. The Control Group F were not asked to

recall the picture, but at time intervals equivalent to those for the Experimental Group they were also given the recognition test of the original photograph. Under these conditions it was hoped that subjects would not anticipate or prepare for a test of recognition. The results are shown in Table III.
The effect of the interpolated experience of recalling the picture is again apparent on subsequent identification of the picture when there is a time interval between recall and identification.

It is important to compare the answers of those who were, and those who were not able to recognize the original after the interpolated experiences. Why, for example, were 2 subjects able to identify the photograph after recalling it and the other 10 unable to do so? How far do these 10 subjects differ in their original recall and what is the relationship of recall to final identification of the original? Table IV shows the number of items correctly and incorrectly recalled by each of the seven subjects who were given the recognition test of the photograph within ½ hour of recall.

TABLE IV

	Items recalled		
	Correctly	*Incorrectly*	*Importations*
Correct identification			
Subject 1	6	0	0
2	6	1	0
Incorrect identification			
Subject 3	4	1	3
4	7	0	2
5	13	1	1
6	7	2	2
7	7	2	1

Importations.

It will be seen from Table IV that the subjects who identified the picture correctly were the only 2 of the group who made no importations. Subject 3 who inserted most erroneous details was the only one of the group who was "Very Sure" that his incorrect response of "No—that is not the original" was correct. Similar results were "obtained from an analysis of the items correctly and incorrectly recalled by the 64 subjects in the Poster Experiment and from a qualitative treatment of their replies, e.g. "No—not this picture—the original had a 3rd car" (*Importation*) where, in fact, the 3rd car had existed only in the interpolated recall.

Dominance of Detail.

Certain items in the display seemed to be more dominant than others to individual subjects and to facilitate and direct both recall and recognition when there was not the interpolated experience, e.g. "Yes—same picture, I remember it looked like a Jaguar car . . . the bicycle lamp was in a peculiar place . . . the boy had a bundle of papers under his arm." On the other hand, when the recognition test was applied after recall it was these originally dominant details which did not now appear to "match"—e.g. "I remember seeing the poster because it looked as if it had been painted by an amateur; the car was badly drawn—it is the kind of car I might have drawn. . . ." In the recognition test this subject said "This is not the same picture —the car looks better drawn; the car figured mainly in my memory and on that

score I am beginning to be more and more sure that it is not the same picture." Another subject who constructed his recall round a "Central Archway" which ". . . looked like the Archway in Downing Street site" said "No—the original was a local scene." It was these dominant details which most clearly lost their identity.

Omissions.

Items which were omitted in the recall often seemed to assist in determining the response in identification, e.g. "No—not the same—the original had no cyclist in it" (i.e. when the subject had failed to recall the cyclist). In contrast to this, if the subject had a recognition test only, he showed great willingness to accept any details which he did not remember having seen in the original, e.g. "I don't remember having seen a cyclist in the original, but I expect he was there."

Related Experiences.

Many of the errors in identification after recall were due to related and relevant items being inserted in the recall—Belisha Beacons, parts of other road safety posters, either pictorial or verbal, etc.

Familiarity.

In no group was there any evidence of recognition depending upon the appearance in consciousness of an ultimate and irreducible feeling of familiarity or unfamiliarity with the display as a whole. In general, the subjects appeared to analyse the detail in order to look for something which was *not* familiar and, in so doing, to remark on the items which were familiar. It was these which seemed to determine recognition when there had been no interpolated recall.

The Effect of the Different Wording in the Recognition Test.

A comparison of the scores of Control Group C with those of Control Group D and those of Groups A and B suggests that substitution of different wording from the original in the recognition test had considerable effect on ultimate recognition of the picture—whether recall was interposed or not. This effect was apparent not only in accuracy but also on certainty of decision in identifying the pictures. Table V shows that there were less subjects in Group D who were able to identify the picture accurately; of these 9 subjects only 4 were Very Sure or Sure of their decision as compared with 11 of the 14 of Group C for whom there had been no change of wording in the recognition test.

TABLE V

	Identification	
	Number of subjects	
	Correct	*Incorrect*
Group C—Recognition test only, Poster X	14	2
Group D—Recognition test only, Poster Y ..	9	7

A comparison of the tables shows that the combined effect of the recall together with the change in detail in the recognition test has reduced the number of correct responses to nil. Of the 16 incorrect responses, 11 were Very Sure or Sure of their decision; 5 were Doubtful.

IV

Discussion

The findings of this experiment are consistent with those of Davis and Sinha (1950) who report that an interpolated experience brings about changes in the recognition of an original experience. In their experiment the original experience was the hearing of a story and the interpolated experience was formed by a picture. The present experiment differs in that the interpolated experience was recall of the original experience. In this way it was similar to the design used by Zangwill (1939) whose subjects reproduced a prose passage before being given a Recognition Test, and by Postman, Jenkins and Postman (1948) on their work with nonsense-syllables. The results of all these experiments suggest that accurate recognition is unlikely after recall.

A comparison with the results obtained by Davis and Sinha on the effects of an interpolated experience on recognition reveals an interesting variation in the quantitative effects of the interpolations. Whereas these experimenters reported that 5 of 20 subjects were influenced by the interpolated experience in identifying the original, in these experiments the effect (except where recall was completely accurate) was marked by its lack of exceptions. The explanation seems most likely to be found in the differences between the tasks. So long as the interpolated activity (where that activity is anything other than recall) may be given a separate identity in memory there must be items which are remembered as referring to that activity alone. Wherever these items occur they would serve as cues to separate the interpolated experience from the original experience and so be liable to prevent "recognition" of anything other than the original material. But when recall is the interpolated experience there are no items which individualize it. Any foreign items will be reabsorbed into the "schema" and there will be no "cues for rejection."

In the present experiment an attempt has been made to depart as far as possible from the restrictive influences of the set task in a laboratory situation by using incidental perception as the original experience and to analyse the factors which militate against the identification after recall. These factors seem to be two-fold:—

(*a*) *Importation of items into the recall.*—From the subjects' introspections it was clear that when recall had been interposed recognition seemed to be determined by the success or failure of finding in the picture those items which had been recalled. Where recall had included erroneous items, it was the absence of those items which now determined incorrect recognition.

(*b*) *Omission of items in the recall.*

Postman, Jenkins and Postman in their work on nonsense syllables state that the accuracy of recognition is lowered through weak associations dropping out during a process of recall. This is only half the answer, however, since their experimental material was not conducive to a consideration of what happened when items were imported in the recall. In the present experiments, where items were imported, they were highly relevant to the general theme of the picture and were evidence of the effect of past related experiences on an attempt to remember. Such processes are essential to recognition and to remembering and would not be apparent from a study of nonsense syllables.

It seems likely from the results of these experiments that the amount of correct detail which a subject is able to recall is less important in determining eventual identification of the original material than the number of items he erroneously

imports in his effort, presumably, to fill the gaps. These inventions seemed to assume dominance. Bartlett states ". . . dominant details are a kind of nucleus about which the rest cluster. They set the stage for remembering." In these experiments the dominance of certain items did set the stage for remembering, but, after recall, "upset" the stage for recognition. The act of identification revealed that the scene-shifters had worked too hard. These dominant details seemed to direct recall or recognition, but they were those which favoured least the preservation in memory of a separate identity of experience and seemed those which merged most into an organization of related experiences.

References

1. Bartlett, F. C. (1932). *Remembering.* Cambridge.
2. Davis, D. R., and Sinha, D. (1950). The influeuce of an interpolated experience upon recognition. *Quart. J. Exp. Psychol.*, **2**, 132–137.
3. Leo Postman, W. O. Jenkins and Dorothy L. Postman (1948). An experimental comparison of active recall and recognition. *Am. J. of Psychol.*, **61**, 511–519.
4. Whitfield, J. W. (1947). Rank correlation between two variables, one of which is ranked, the other dichotomous. *Biometrika*, **34**, 292–296.
5. Zangwill, O. L. (1939). Some relations between reproducing and recognizing prose material. *Brit. J. Psychol.*, **29**, 371–82.

(*Received 21st July*, 1950.)

AN UNUSUAL CASE OF AFTER-IMAGERY FOLLOWING FIXATION OF AN "IMAGINARY" VISUAL PATTERN

BY

L. WEISKRANTZ*

(From the Institute of Experimental Psychology, Oxford)

The case of a healthy normal subject is described in which negative after-images were reported to follow fixation of an "imaginary" pattern. The size of the after-image was measured at various distances and found to agree closely with the size of an after-image induced by fixation of a "real" visual pattern, and to conform roughly with Emmert's Law. No evidence for chromatic after-images was obtained. Two controls were conducted, one indicating that the results could not be attributed to after-images resulting from objective aids used to maintain an "imaginary" pattern of constant size; the other discounting the possibility of the subject's having guessed the correct values. Similar phenomena reported in the early literature are briefly mentioned, and some theoretical implications discussed.

I

Introduction

In a study quite incidental to the present observations, it was necessary for subjects to "imagine" that a black square was located on an objectively blank piece of white cardboard with which they were presented, and to fixate a point on the contour of the "imaginary" square for at least a minute. Needless to say, it proved very difficult for most randomly selected subjects to accomplish such a task to their own satisfaction, although a few subjects claimed to have achieved it. One subject, in particular, Mrs. A.B., aged 24, reported that she could carry out the feat of imagination with little difficulty, and startled the experimenter by remarking, after having carried out the required fixation, that she then had a clear negative after-image of the square. The fixation was repeated several times over a period of several days with similar results.

Similar cases are not entirely unknown. Woodworth (1938, p. 561) refers to the "cortical after-image in 'memory after-images'," although it is not entirely certain that he is referring to the phenomenon under consideration here, particularly since the phrase "memory after-image" has sometimes referred in the past to eidetic imagery (Allport, 1924, p. 100). William James (1896, p. 66) refers to a paper by G. H. Meyer, published in 1843, in which Meyer claims: "With much practice, I have succeeded in making it possible for me to call up subjective visual sensations at will. . . . I can now call up before my eyes almost any object which I please. . . . Most of these subjective appearances, especially when they were bright, left after-images behind them when the eyes were quickly opened during their presence. For example, I thought of a silver stirrup, and after I had looked at it a while I opened my eyes and for a long while afterwards saw its after-image" (Meyer, 1843, p. 233). James comments that "The negative after-images which succeeded upon Meyer's imagination . . . are a highly interesting, though rare, phenomenon. So far as I know there is only one other published report of a similar experience"—that of Ch. Féré (1885, p. 364). Johannes Müller (1826, p. 36) quotes an account by Gruithuisen (1812)

* Working with a Catherwood Foundation Fellowship.

in which after-images in a waking state were reported to have followed vivid imagery in a dream:

"Mir träumte ich zeige einer Dame die schön violettblaue Farbe des Flusspathes auf glühenden Kohlen. Diess Experiment gelang im Traum scheinbar so gut, dass mir davon die Augen wie im Sonnenlichte geglendet wurden. Darüber erweckte ich mich, und ich hatte im Auge einen gelben Fleck. Dieser Fleck wurde endlich violettschwarz, dann öffnete ich die Augen, da war er gegen das Fenster gehalten, dunkler als die anderen Stellen des Auges und bewegte sich genau wie andere Täuschungen in Wachen mit den Augen über die Gegenstände hin." (p. 256.)

Further interest in these rare cases seems to have died out with the passing of anecdotal psychology and with the eclipse of speculation about central factors in the origin of after-images by the rapidly advancing knowledge of retinal adaptation processes. As far as the present writer is aware, there have been no cases reported in the literature since James drew attention to Meyer and Féré. If, however, such a phenomenon can reasonably be held to be genuine, it assumes some importance, as James was careful to point out (1891, pp. 70–71).

Unfortunately, Mrs. A.B. is well acquainted with after-image phenomena, having studied psychology at Rhodes University College and being at present employed at the Institute of Experimental Psychology. While there was no reason to suppose that she was not giving a genuine phenomenological report, her knowledge combined with her quite vivid powers of imagination might have convinced her that she saw what she would normally have expected to see on comparable inspection of a real visual pattern. The following observations of the size of the after-image with varying distance, however, make it difficult to believe that the phenomenon was not genuine in a more fundamental sense.

It is already well known that the linear dimensions of a negative after-image vary approximately linearly with the apparent distance at which the after-image is seen. The relationship, known as Emmert's Law, can be stated as follows (Boring, 1942, p. 292): $S_1/D_1 = S_2/D_2$ where S indicates linear dimensions of size and D distance from the observer. The purpose of the investigation was first to determine whether the size of the after-images reported by Mrs. A.B. followed this law, and if so how closely; and second, to determine whether, and if so to what extent, there was a correspondence in size between the "imaginary" after-images and those following fixation of a real visual pattern.

II

Methods and Results

Four faint black dots were marked on a large white sheet of cardboard which determined the corners of the "imaginary" black square; the square measured 2·25 in. × 2·25 in. A fixation point was also marked so that it fell on the left vertical contour of the square. The subject always sat at a distance of 63 in. from the cardboard. To determine the size of the after-image, a large sheet of 0·1 in. graph paper, with a fixation point marked in its centre, was placed at various distances from the subject, the experimenter adjusting a pair of dividers just in front of the paper until the subject expressed satisfaction that the dividers indicated the proper size. It was thereby possible to make a quick determination of the size of the after-image to the nearest 0·1 in.

In a second series, the procedure was identical, except that instead of fixation of an "imaginary" square determined by four dots, the subject was presented with a white cardboard actually containing a black square of the same dimensions as the

"imaginary" square (2·25 in. × 2·25 in.). The same technique for measuring the size of the after-image was used. The "imaginary" series never consisted of more than two judgments at any one sitting, since the subject found the "imaginary" square difficult to sustain for long periods. The whole of the "real" series was conducted at one sitting after the "imaginary" series had been completed. The order in which the distances were presented was randomized in both series. The results obtained in the two series together with the corresponding predictions based on Emmert's Law are those in Table I.

TABLE I

Distance (in.)	*Size of after-image (in.)*		*Predicted size (in.)*
	"Imaginary" series	*"Real" series*	
178·00	7·3	7·3	6·4
132·50	5·0	5·0	4·7
122·00	4·5	4·6	4·4
41·50	1·8	1·4	1·5
40·00	1·3	1·3	1·4
33·00	1·8	1·1	1·2

It can be seen at once that there is remarkably close agreement between the "imaginary" series and the "real" series. Four of the six pairs of readings fall within 0·1 in. of each other. Although the subject knew vaguely that the after-images should grow at greater distances and shrink at shorter distances, it seems most unlikely that she could have guessed the values with such accuracy. In this connection, it must be borne in mind that the "real" series was not conducted until the "imaginary" series had been completed and that the random orders were different in the two series. In addition, it will be noted that the sizes of both the "real" and the "imaginary" series agree roughly with predictions based on Emmert's Law. This conforms to the observation of Young (1948) that perfect conformity with Emmert's Law is exceptional.

It could be objected that the subject might have obtained an after-image of *only* the four faint dots in the "imaginary" series and "filled in" the area to form a solid, square after-image. This possibility appeared unlikely in view of the faintness and smallness of the dots, as well as the difficulty of seeing such tiny after-images on closely-ruled graph paper. Nevertheless, a test was conducted in which Mrs. A.B. was asked to look at the pattern containing the small dots, without imagining that they formed a solid square. Afterwards she was shown a blank white card with a fixation point upon it, and asked to report any after-images. She replied that, to her own surprise, she could not see even the faintest after-image of the dots. This objection, therefore, is unlikely to be valid.

Another investigation was made to determine whether a highly sophisticated subject could *guess* the size of after-images projected at various distances on the basis of Emmert's Law. The subject, a graduate student of psychology, with qualifications in physics, was well acquainted with Emmert's Law; the same test cards were employed. The subject was shown the card with the four faint dots for 5 seconds and then asked to guess the "Emmert's Law size" of the after-image on the graph paper located at various distances. He was able to do this with some confidence within 5 or 10 seconds. All readings were conducted at the same sitting, and the

order of presentation was randomized. The sizes at the same distances as are listed in Table I were, respectively, 2·8 in., 2·6 in., 3·0, 2·0, 2·0, and 2·2. It can be seen at once that the values diverge very widely from Emmert's Law and are not distributed linearly. It seems unlikely, therefore, that the results given by Mrs. A.B. could be due merely to sophisticated guesswork.

So far we have been dealing only with achromatic after-images. It was difficult to test convincingly for the presence of chromatic after-images since the subject was well acquainted with most of the common chromatic complementaries. According to the subject's own account, however, she did not obtain chromatic after-images following prolonged fixation of an "imaginary" coloured square. One test was conducted which was intended to make it difficult for the subject to make quick use of her knowledge of chromatic after-images. She was instructed to fixate an "imaginary" blue square. Afterwards she was shown two green squares with a fixation point between them. If any chromatic after-image appeared, it is plain that the hue of one of the squares would be affected. The subject was asked to report quickly whether the two squares appeared different in hue. Her reply was that there was no perceivable difference in hue, although the critical square appeared a bit brighter. This observation supports the subject's claim that she did not experience chromatic after-images, but only achromatic after-images as "imaginary" after-effects.

It was thought that Mrs. A.B.'s unusual ability might be associated with eidetic imagery. A test for eidetic imagery, however, using Allport's "Gartenwirthschaft" picture (1924, reproduced in Boring, Langfeld, and Weld, 1939, p. 365), yielded negative results. After the picture was shown to the subject for 35 seconds, she reported that she could see no image of the picture when she looked at a blank white screen. She could, however, list several details with a fair degree of accuracy. Nevertheless, her reproduction of the picture one minute after the inspection indicated that she had recalled the general structure in the usual way and there was no good evidence that she had conjured up an exceptionally accurate or complete image. Thus she could remember that there was a "German sign," but could not say what the word was. It appears, then, that her unusual visualizing ability is not eidetic in the conventional sense of the term.

The subject was questioned as to whether she had ever noted after-images of "imaginary" patterns before taking part in the above-mentioned experiment, and she replied in the negative. But she reported that she had the "usual" after-image experiences. For example, if she reads a book in sunlight, she habitually experiences very strong negative after-images of the material. She dislikes looking at bright red-and-white striped candy in a brightly lit shop window as she gets "green and black" after-images. Under normal indoor conditions of illumination, however, she does not get strong or lasting after-images. While she was of the opinion that such after-images are quite common with most individuals, it appears that she is rather more sensitive in this respect than the average person.

Certain other incidental observations should be noted. The subject reported that the size of the after-images in the "imaginary" series oscillated slightly for the first few seconds, but then became relatively constant. Such oscillation was not noted in the "real" series. The "imaginary" after-images were much less vivid than the "real" ones, and generally lasted only about five or six seconds, although in one instance the subject reported a negative after-image in the "imaginary" series 55 seconds after the end of the fixation period. The ability to perform the necessary inspection of an "imaginary" pattern seemed to vary markedly and unsystematically

from day to day; on some days it was virtually impossible for her to perform to her own satisfaction. Very bright conditions of illumination made this type of visualizing much more difficult but conditions of normal illumination appeared entirely satisfactory.

III

Comment

These observations are of interest in at least two respects. In the first place, as has already been mentioned, individuals who report negative after-images following fixation of an "imaginary" pattern appear to be exceedingly rare. It is quite likely, however, that a well-directed inquiry would show that such cases are more common than is generally supposed. In the second place, this phenomenon, apart from its rareness, has important implications for any theory of after-imagery. Theoretical advance often comes about by consideration of the esoteric and the unusual as well as the normal and well-known. The present observations are difficult to explain on any theory of after-imagery which places exclusive emphasis on peripheral factors. Unless the unlikely assumption is made that Mrs. A.B.'s visualization of the "imaginary" square induced concomitant retinal changes, only changes in central processes can explain the after-images which she reported. James, in his discussion of these phenomena, does somewhat unconvincingly conclude that "the current demonstrably *does* flow backward down the optic nerve in Meyer's and Féré's negative after-images (cf. above). Therefore it *can* flow backwards; therefore it *may* flow backward in some, however, slight degree in all imagination" (1891, pp. 70–71). This conclusion would appear, on first blush, to contradict the evidence adduced by Craik (1940) that after-images may be obtained even after the optic tract has been rendered temporarily inactive. Actually, however, the dilemma can be resolved by assuming that both a peripheral and a central factor are normally involved in achromatic after-images, but that under certain circumstances either factor may operate as a sufficient condition. This would account both for the present results and for those of Craik. Further, it would fit in nicely with the theory of after-imagery which has been presented by Woodworth (1938, pp. 558–561), based on a modification of Ebbecke's views (1929), and which includes a consideration of both central and peripheral variables. The view that factors of central origin may be involved in after-images is also supported by the allegation that after-images are obtained following fixation of hallucinatory patterns induced under hypnosis (Binet and Féré, 1887, p. 188; Erickson and Erickson, 1938). One might even suggest very tentatively that the cortical aspect of after-image processes might well be part and parcel of figural after-effects, reported by Köhler and Wallach (1944), although it has been implicitly assumed, up to the present time, that the two phenomena are quite disparate.

The writer wishes to thank Prof. G. Humphrey for facilities in the Institute of Expefimental Psychology, Oxford, and to express his appreciation to Dr. R. S. Creed for discussing this paper with him and providing historical references. Mr. O. L. Zangwill and Mr. R. C. Oldfield have offered very helpful advice.

References

1. Allport, G. W. (1924). Eidetic imagery. *Brit. J. Psychol.*, **15**, 99–120.
2. Binet, A., and Féré, Ch. (1887). *Le magnétisme animal.* Paris.
3. Boring, E. G. (1942). *Sensation and perception in the history of experimental psychology.* New York.

4. Boring E. G., Langfeld, H., and Weld, H. (1939). *Introduction to psychology.* New York.
5. Craik, K. (1940). Origin of visual after-images. *Nature*, **145**, 512.
6. Ebbecke, U. (1929). Uber positive und negative Nachbilder, ihre gegenseitige Beziehung und den Einfluss der lokalen Adaptation. *Arch. f. d. ges. Physiol.*, **221**, 160–212.
7. Erickson, M. H., and Erickson, E. M. (1938). The hypnotic induction of hallucinatory color vision followed by pseudo-negative after-images. *J. exp. Psychol.*, **22**, 581–588.
8. Féré, Ch. (1885). Sensation et mouvement. *Rev. Phil.*, **20**, 337–368.
9. Gruithuisen (1812). *Beiträge zur Physiognosie und Eautognosie.* München.
10. James, W. (1891). *Principles of psychology*, Vol. 2. London.
11. Kohler, W., and Wallach, H. (1944). Figural after-effects. An investigation of visual processes. *Proc. Amer. phil. soc.*, **88**, 4, 269–357.
12. Meyer, G. H. (1843). *Untersuchen üb. d. Physiol. d. Nervenfaser.*
13. Müller, J. (1926). *Uber die phantastischen Gesichtserscheinungen. Eine physiologische Untersuchung.* Coblenz.
14. Woodworth, R. S. (1938). *Experimental psychology.* New York.
15. Young, F. A. (1948). The projection of after-images and Emmert's Law. *J. gen. Psychol.*, **39**, 161–166.

(*Received 5th July*, 1950.)

THE RELATION BETWEEN ANGLE OF DISPLAY AND PERFORMANCE IN A PROLONGED VISUAL TASK

BY

D. C. FRASER

(*From the Psychological Laboratory, Cambridge*)

An experiment is described to test the relation between angle of display and performance in prolonged visual tasks, using the Clock Test in three positions of display surface, vertical, at an angle of 45 degrees to the horizontal, and horizontal. In every case the line of regard was kept at right-angles to the display surface. The results indicate that significantly fewer stimuli are missed in the vertical position of the display surface than in the other two.

The deterioration observed in the second half-hour of the original Clock Test experiments was not found in the present experiment, and two possible reasons for this are discussed briefly.

I

Introduction

While the writer and Dr. W. E. Hick were working on some experiments in plotting (Hick and Fraser, unpublished), the problem was raised as to the best position to mount the viewing screen of cathode-ray tube and similar displays "so as to cause minimum fatigue to the operator." Now the term "fatigue" is usually regarded with a good deal of well-merited suspicion by psychologists, and the difficulty of obtaining a reliable, a consistent, or even a meaningful measure of fatigue is considerable. It was felt at first that the criterion might have to be a subjective one, but, after discussion and rejection of several alternative designs, it occurred to the writer that use could be made of Mackworth's Clock Test, employed in some previous experiments on prolonged visual search.

Mackworth himself preferred to call the phenomenon studied "breakdown of vigilance," or "deterioration in performance," but if we accept Thorndike's definition of fatigue as "that loss of efficiency resulting from continuous work which can be eliminated by rest" (Thorndike, 1912), we can perhaps take the decrement of performance measured in the Clock Test as a satisfactory criterion. The present experiment was, accordingly, undertaken to discover if any significant difference could be obtained in the pattern of deterioration using the Clock Test in three positions of display surface, vertical, at an angle of 45 degrees to the horizontal, and horizontal. In every case the line of regard was kept at right-angles to the display surface.

A full description of the display and procedure may be found in Mackworth (1950). Briefly, the subject sits looking at a black pointer 6 inches in length which normally moves like the second-hand of a large clock in front of a vertical white surface. Every second it jerks on to a new position, one hundred of these movements making up the full circle. There are no scale markings or reference points of any kind on the background. The subject is told beforehand that every now and again at long and rather irregular intervals the pointer will move twice the usual distance, giving the effect of an unusually large jump forward. He is to watch for these movements, and press a response key as soon as he notices one.

Under these conditions, Mackworth found that there was a significant difference between the average performance on the first half-hour of the test and that on the second, third or fourth half-hour.

II

Procedure

The administration of the Clock Test in this experiment differed in several respects from the procedure in Mackworth's experiments:—

The clock was placed at a distance of 1½–2 feet from the observer to correspond roughly to the position of an operator watching a cathode-ray tube display. As in the original clock experiments, the observer was placed 7 feet away from the clock; this meant that the double movement was correspondingly greater, although the relation between the single and double movements remained the same.

In the original experiments the subject was in a closed cabinet by himself, and as far as possible all stimuli which might distract or rouse him to alertness were excluded. In the present experiment the writer sat in the same room with the subject, and there was no particular attempt to eliminate the ordinary external street noises, etc. Under operational conditions it is to be expected that the observer will be subject to a random selection of environmental stimuli, and it is reasonable to suppose that the effects of these would tend to cancel each other out over a long period.

To test whether the Clock Test would give useful results under the modified conditions, preliminary investigations were carried out on a number of subjects. These indicated that the differences mentioned above were not likely to invalidate the procedure, so the experiment proper was undertaken. One modification worthy of note was to substitute a matt white card for the glossy xylonite surface of the clock face, after complaints from the subjects about reflections from this surface.

The subject sat in a straight-backed chair, chosen so that he could sit upright with reasonable comfort but without any inducement to become too relaxed. The clock was placed so that the face was perpendicular to the subject's line of sight in each position. To achieve this, the subject sat upright for the vertical position, had to lean forward a little and tilt his head for the 45° position, and had to lean forward to a marked extent for the horizontal position.

Seventeen subjects were tested under the final arrangement of conditions. Each subject was tested in each of the three positions, in random order, each test lasting for one hour, and the usual precautions were taken to combat increasing boredom and irritation by assuring the subject that "the results were valuable," etc. A practice period of five minutes was given before each test spell. An experiment on the Clock Test by Carpenter (1946) has shown that a preliminary practice period of five minutes is adequate, and that there is no appreciable learning of the sequence even up to six repetitions of two hours each with the same subject.

III

Results

The full results for the seventeen subjects are given in Table I, and the Analysis of Variance in Table II. The total of missed stimuli (errors) for the vertical display position is 35, for the 45 degree position 62, and for the horizontal position 79, giving a mean error score of 2·059, 3·647 and 4·647 for the vertical, 45 degree and horizontal positions respectively of the display surface. The mean error score for the vertical position compares reasonably closely with the mean error scores in the previous Clock Test experiments, considering the changes in the experimental procedure.

TABLE I

TABLE OF RESULTS

Preferred position	*Subject*	*Missed stimuli (errors) in each position*		
		Vertical	45°	*Horizontal*
H.	1	1	5	3
H.	2	7	6	17
45°	3	0	0	2
45°	4	1	8	8
V.	5	2	2	2
45°	6	2	3	2
45°	7	2	9	8
V.	8	1	9	5
V.	9	3	4	6
45°	10	0	4	3
N.P.	11	0	2	1
V.	12	1	3	5
V.	13	2	1	1
45°	14	2	0	2
V.	15	4	4	3
45°	16	6	1	9
N.P.	17	1	1	2
		Total 35	62	79
		Mean 2·059	3·647	4·647

Critical ratios, for *Vertical*—45°	2·75	Significant
for *Vertical*—*Horizontal* ..	4·48	Significant
for 45°—*Horizontal*	1·73	Not significant

The Analysis of Variance table shows that the Variance Ratio (F) of Positions to Remainder is 5·11. With two and thirty-two degrees of freedom respectively, this result is significant at considerably better than the 5 per cent. level, in fact closely approaching the 1 per cent. level. It will be observed that the distribution is rather skew, and the results at first sight seem to be heavily influenced by the unusually poor performance of Subject No. 2. It is, therefore, reassuring to note that a repetition of the Analysis of Variance with sixteen subjects, omitting Subject 2, slightly increases the F ratio.

TABLE II

ANALYSIS OF VARIANCE

	Sum of squares	*D.F.*	*Mean sq.*	*F.*	5 *per cent. pt.*	1 *per cent. pt.*
Total	524·63	50				
Positions ..	57·92	2	28·96	5·11	3·3	5·36
Subjects ..	285·3	16	17·83	3·14	1·97	2·62
Remainder ..	181·41	32	5·67			
Repetition omitting Subject No. 2						
Total	313·92	47				
Positions ..	48·98	2	24·49	6·31	3·3	5·4
Subjects ..	148·59	15	9·906	2·55	2·02	2·7
Remainder ..	116·35	30	3·378			

The Standard Error of Error Means is 0·577. Taking the ratio of the difference of means for each pair of positions to this S.E., we find that the critical ratios are significant for (V–45) and (V–H), but not for (45–H), i.e. the mean of V differs significantly from those of 45 and H, but the latter two are not significantly different. (Significance here is taken as a Critical Ratio greater than 2·5, a reasonably exacting standard.)

Normally, results in the Clock Test are also considered from the point of view of the incidence of slow responses. Unfortunately, in this case some of the response times are suspect owing to a slight defect which developed in the recording apparatus. There was some suggestion of a tendency to increased response times in the horizontal position as compared with the other positions, but the data are perhaps too inexact to justify further analysis.

A period of one hour was chosen as the unit here partly because it was expected that this would be the maximum time which an operator would be working on this type of vigilance task, and partly because Mackworth's experiments showed that we could expect a significant increase in the number of errors made in the second half-hour period as compared with the first. Table III, which gives the total number of errors for all subjects in each of the two half-hour periods for the three positions, indicates that this has not happened. There is no significant difference between the errors occurring in each half-hour period in any of the positions, and any difference is in favour of the second half-hour period. The consistency of the ratio between the errors made in each period for the three positions is quite remarkable.

TABLE III

Comparison of Errors Made in the First and Second Periods

Position	*Missed stimuli (errors)*	
	First half-hour	*Second half-hour*
Vertical	19	16
45 degrees ..	33	29
Horizontal ..	41	38

IV

Discussion

The statistical picture is clear, but there remains the problem of interpreting it, both in terms of its relation to operational requirements and its relevance to psychological methodology.

Clearly, the experiment has yielded a definite result, and one of value, if it is legitimate to generalize from it to other vigilance tasks. It seems reasonable to conclude that a vertical or near-vertical display is to be preferred where the subject has to watch a display for a considerable time. There is, of course, the possibility that the results are specific to the present experimental situation, and might be altered by a different type of chair, display or task, or that the results are due to a chance error of sampling. The consistency of the error scores from the first to the second half-hours, and the general trend of the results seem to be against this, however.

The experiment has probably answered the original problem raised, since the layman's conception of "fatigue" is very general and covers almost any decrement in performance (cf. Robinson, 1934). But it is doubtful if we can call the effect observed by us a "fatigue" effect, even in the Thorndikian sense, since there was no evidence of a decline in efficiency due to prolongation of work.

It is interesting to speculate on the reason for this difference from Mackworth's findings. It may, of course, be simply due to the change in the stimulus magnitude. In the present experiment, this is almost four times as great as in the original Clock Test experiments, and it might be expected that the subject would be able to perceive the double movements more easily, and that his performance would not fall off in the same way. On the other hand, one of Mackworth's experiments was undertaken to determine the effect of reducing the length of the clock hand, and therefore the relative stimulus magnitude, by half; and this alteration did not significantly affect the performance. It seems as if the subjects are reacting to a relationship between the single and double stimuli rather than to the physical magnitude of the stimuli. Against this, the percentage of missed stimuli in the first half-hour when the vertical position was employed in the present experiment is lower than in the original Clock Test experiment (9·3 per cent. as against 13 per cent. to 15·7 per cent.), which seems to indicate that the subjects found the present set-up easier.

If the difference between the findings is not due to the alteration in the stimulus magnitude, there is the possibility that it depends on the environmental factors. In one case the subject is entirely alone, and as far as possible all external noise and stimuli are excluded. In the other, the experimenter is in the room with the subject, and the latter is treated to a random sample of environmental stimuli. It would be too much to generalize from the results of one experiment, but if the typical deterioration pattern found by Mackworth is not repeated in other vigilance tests, it should be worth while considering how far the environmental factors influence the performance.

The apparent superiority of the vertical position is not easy to explain on general psychological principles although it is not difficult to give a *post hoc* explanation by selecting appropriate generalizations. For example, we can postulate that the vertical position is the one which involves the minimum of muscular strain, since the head and upper part of the body have to be tilted in the other two positions; we might therefore expect that the subject's performance would be less liable to fall off in the vertical position. Or we might take it that the inclined display position is one more favourable to the relaxation of attention; in other words, that the subject is compelled to remain more alert in the vertical position. It is a well established fact that a certain degree of strain or tension may favour performance in many types of skill.

In an endeavour to throw some light on this aspect of the problem, a fourth position was arranged in which the clock was mounted above the subject's head at an angle of 45 degrees, so that he had to tilt his head slightly backwards. (The writer is indebted to Dr. W. E. Hick for this suggestion.) The first subject to be tried out in this position, however, complained after a short time of headache and dizziness, and the trial had to be terminated. The same thing occurred with the next subject; and it was decided to abandon this position. The effect of this is mentioned, however, since in some ships certain instrument dials, etc., are mounted so that the observer has to tilt his head upwards to observe them.

It will be noted that the subjects' preferences are given in the first column of Table I. The subject was encouraged to give his introspections as fully as possible,

but as in general there appeared to be a marked preference for one position only, no attempt has been made to analyse the preferences. These are almost equally divided between the vertical and 45 degree positions, with two subjects choosing the horizontal position. The introspections were, on the whole, vague and unhelpful, and a comparison of the preferred position with the subject's actual performance serves to indicate how unreliable a purely subjective criterion might have been.

My thanks are due to Dr. Carpenter and Dr. Hick for discussion and the loan of papers and apparatus, and to Mr. Chambers for checking the statistics.

References

1. Carpenter, A. (1946). An experiment to determine whether the performance on the Clock Test improves with practice. Applied Psychology Research Unit Report No. 41 (unpublished).
2. Mackworth, N. H. (1950). Researches on the measurement of human performance. Medical Research Council Special Report Series, Report No. 268. H.M. Stationery Office: London.
3. Robinson, E. S. (1934). Work of the integrated organism. In Handbook of General Experimental Psychology, Murchison, C. (Ed.), p. 579.
4. Thorndike, E. L. (1912). The Curve of Work. *Psych. Rev.* **19**, 165–94.

(*Received 6th August*, 1950.)

THE EFFECTS OF REPEATED RETESTING: III. FURTHER EXPERIMENTS AND GENERAL CONCLUSIONS

BY

V. R. CANE and A. W. HEIM

STATISTICAL APPENDIX

BY

V. R. CANE

(*From the Medical Research Council Applied Psychology Research Unit, Cambridge*)

This is the third of a series of papers dealing with the effects of repeated retesting on intelligence test scores. It comprises an account of two further experiments, and a discussion of the four so far performed, since each throws light on the results of the others and it is their joint interpretation which is thought to be of value.

The two earlier experiments consisted in repeated testing of a group of W.E.A. students and a group of mentally defective boys by an intelligence test (AH 4) designed for an unselected population. In the two later experiments, a group of Technical College students and a group of Naval Ratings were repeatedly tested on another intelligence test (AH 5), designed for a population of high-grade intelligence, such as university students. Examination of all the results confirms the artificiality of the "ceiling effect" which was suspected in the first experiment; it suggests that, given sufficient scope, subjects of all levels of intelligence would, with repeated trials on the same test, continue to improve their score up to and beyond 10 testings; and that, *given comparable opportunity*, subjects with initially higher scores will tend to improve more than those with initially lower scores. All these results concern short-term experiments with testing at weekly intervals.

The four sets of results considered jointly indicate that conclusions on these problems should be drawn only in terms of the relation between the level of the test and that of the group tested: this offers some resolution of the conflicting findings of earlier investigators.

I

Introduction

In two previous papers (Heim and Wallace, 1949, 1950), the results of repeatedly retesting a group of above average intelligence and a group of high-grade mental defectives on the same intelligence test were described. AH 4, the test used, was originally devised for use with unselected populations and it was thought that some of the results might be due to its being too easy, after a few testings, for some members of the intelligent group. The results yielded by the less intelligent group supported this hypothesis. The average score of the mental defectives increased almost linearly throughout the 10 testings, whereas the average score of the intelligent subjects had shown a progressively smaller increase at each testing: in fact there was little difference between their scores for the 5th and the 8th trials.

In view of these findings two more experiments were performed, again with a group of high and a group of lower intelligence, but on a much more difficult test, AH 5. The intelligent group consisted of students attending evening classes at the Cambridgeshire Technical College and the less intelligent group consisted of Naval Ratings. It was hoped that the Technical College students would prove roughly

comparable to the W.E.A. students (as indeed they did, see Appendix, Figure 5). The Naval Ratings were not of course considered comparable with the mentally defective schoolboys but their relation to AH 5 was thought to be in some ways analogous to the relation of these schoolboys to AH 4.

The AH 5 experiments were undertaken then in the hope of providing some general conclusions on the extent and rate of improvement with retesting; on relative gain made by the initially higher and initially lower scorers and, associated with this, the greater or lesser variance shown by groups after repeated retesting; on individual variation within the group; and on the effects of "bias" at different levels of general intelligence.

The conclusions of previous investigators in the field have been divergent. For example, on the question as to who benefits most from practice: Thorndike (1922) and Vernon (1938) found that the duller subjects tended to improve more than the intelligent ones; Rodger (1936) found that his brighter subjects improved to a greater extent than his average and dull subjects; whilst McIntosh (1944) and McRae (1942) found no relationship between practice effect and level of ability.

On the other questions again, where the data are comparable, the conclusions drawn are conflicting (Dunlap and Snyder (1920), Richardson and Robinson (1921), Dawson (1936), Dave (1938) and Woodrow (1938)). The one point of agreement among all investigators is the fact of a considerable measure of improvement in test performance, with practice either on the same test or on parallel tests.

II

Material

The less intelligent group taking AH 5 comprised 8 Naval Ratings who were available as subjects for a period of 8 weeks. They were thus able to take the test 9 times, at weekly intervals. Their ages ranged from 18 to 20 years and, with 2 exceptions, they had left school at 14. (The 2 exceptions, Nos. 3 and 1 on Figure 2, had left at 15 and 17 years, respectively.) The previous jobs of these Naval Ratings included such varied work as plumber's mate, bank clerk and farmer.

The more intelligent group taking Test AH 5 was composed of 18 students from the Cambridgeshire Technical College evening classes. These were self-selected members of a group of about 25, invited by the Technical College staff to take part in the experiment. Fifteen of the 18 took the test once a week for the complete period of 10 weeks and the Technical College results which follow are based on these 15 subjects. (Their initial scores did not differ in any uniform way from those of the other 3 "starters," who were tested respectively 7, 8 and 9 times.) The Technical College students' age range was 18 to 45 years, with a mean of 25 years. The occupations in which they were engaged during the day included salesmanship, printing, clerical work and household duties.

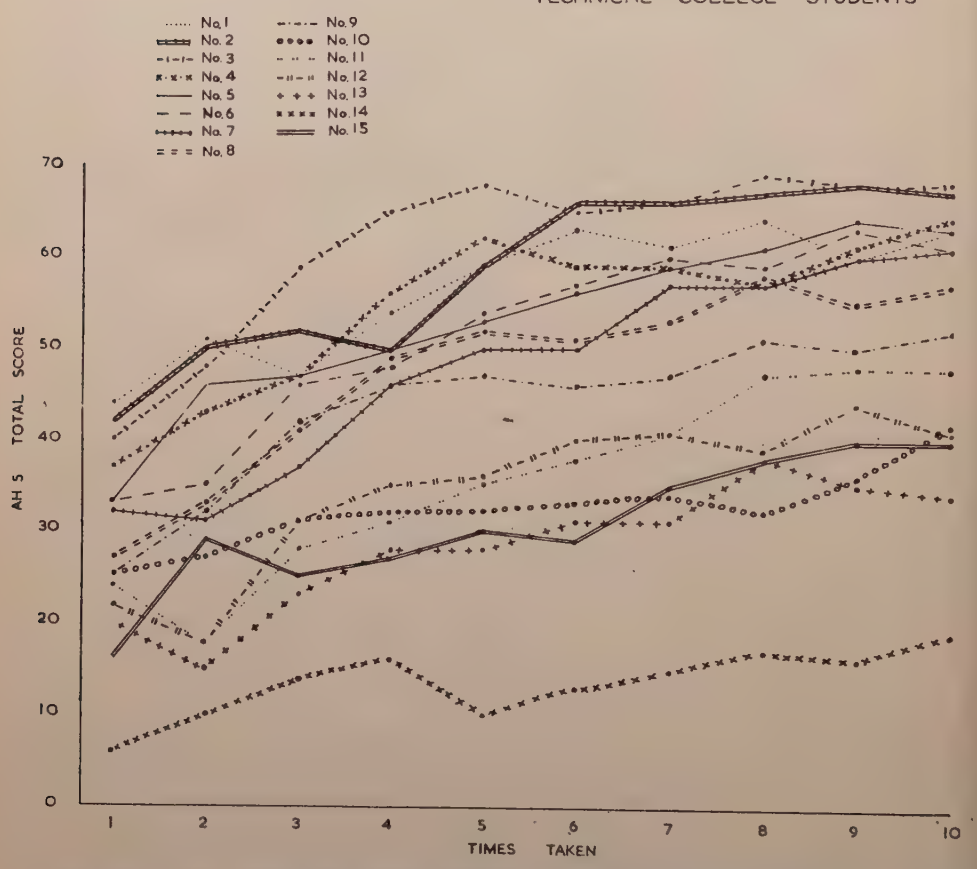
The students, then, were volunteers. They embarked on the experiment because they were interested in it or in psychology generally, and those who stayed the course did so largely out of loyalty to a project to which they had offered their support. "Incentives" for these subjects, suggested by themselves, were a visit to the Psychological Laboratory, a lecture on Psychology and assurance that test results would be communicated to such individual subjects as desired it, at the end of the experiment. No results were divulged in the course of the enquiry.

The Naval Ratings were clearly less interested throughout: they were not volunteers, they found the test difficult and tedious and most of them saw little point in the

experiment. Their main incentive was a "prize" of 10 cigarettes, presented each week (except the first) to the subject who showed the greatest weekly improvement. It was hoped that this expedient would make it unlikely that the *same* Naval Rating would win the prize week after week. Sure enough, 7 different ratings had won their packet by the beginning of the 9th (and last) testing, so the 8th subject (No. 7 on Figure 2) was, on this occasion, awarded a "prize for consistency"!

FIG. 1

INDIVIDUAL PROGRESS CURVES

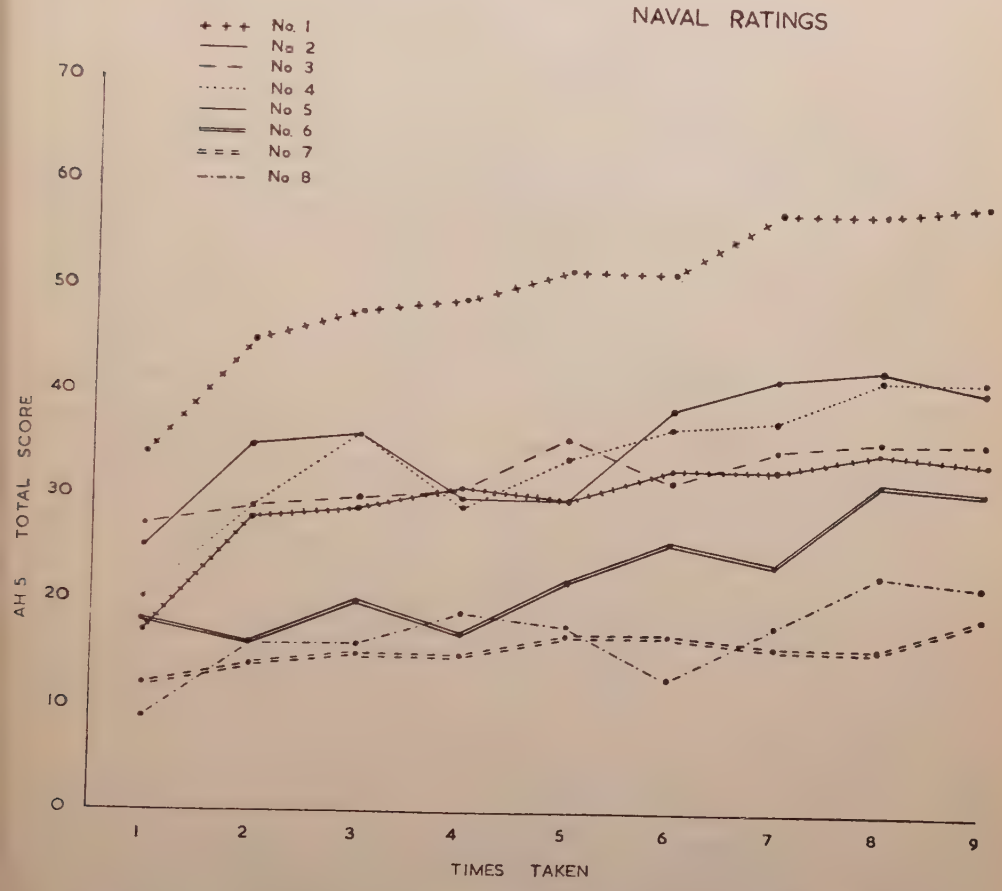


Test AH 5 is an intelligence test intended for preselected subjects of high intelligence (Heim, 1947). It is in 2 Parts, each of which takes 20 minutes and comprises 36 questions. Thus, AH 5 contains approximately half the number of questions in AH 4 and allows twice as long for their solution. As in AH 4, Part I has a verbal and numerical bias and Part II has a visual bias. ("Visual" here signifies reasoning in terms of relations between seen shapes.) Preliminary examples are provided for the subjects to work through at leisure, and complete correctly, before embarking on the test proper. They are allowed to take the questions within each Part, in any order they like; in practice most people take the questions as they are presented, i.e. in roughly ascending order of difficulty within types of question. One marked exception

to this tendency was found among the Technical College students: many of these chose at the 2nd testing to begin at the end and work systematically backwards, because they had not reached the last questions at the 1st session and were interested to see what these questions were like. For this reason, some of the results obtained on that occasion do not fall in with the general trends (see, especially, Figures 1 and 3). When an effect is observable in every case but this, we shall disregard this exception.

FIG. 2

INDIVIDUAL PROGRESS CURVES



In the past, Test AH 5 has been used mainly with populations of university students, lecturers and research workers (Wallace, 1949) and Grammar School children (Harris, 1947–9). It was thought that the questions would present an appropriate degree of difficulty for the Technical College subjects but might prove somewhat too hard for the majority of the Naval Ratings, at least at the 1st session. The test was chosen in this instance on account of the difficulty of the problems and the generosity of the time limit. The writers hoped that these factors would, jointly, reduce the "artificial" ceiling effect which is apt to occur when intelligent subjects are repeatedly retested. Moreover AH 5 repeated produces conditions very near to

AH 4

M.D. SCHOOLBOYS

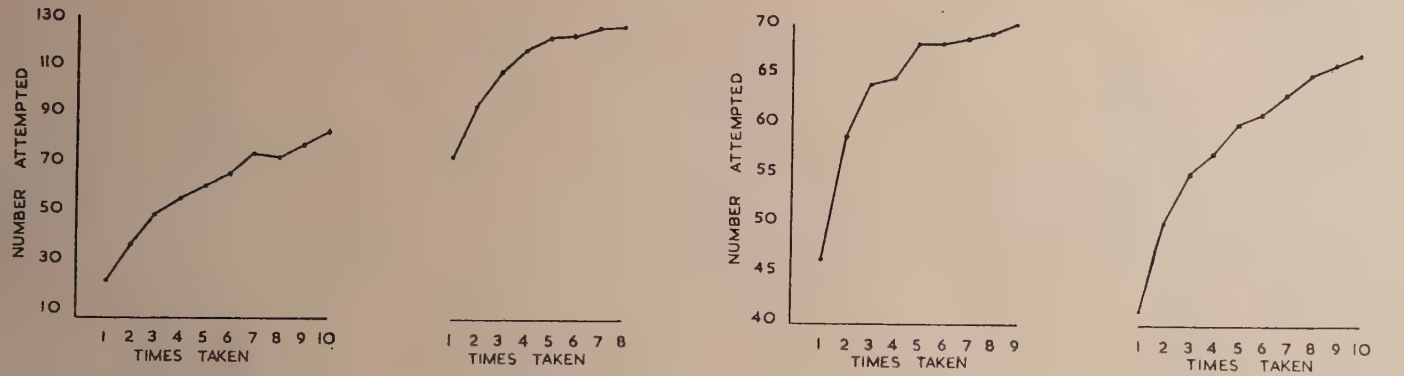
W.E.A. STUDENTS

AH 5

NAVAL RATINGS

TECHNICAL COLLEGE STUDENTS

NUMBER OF QUESTIONS ATTEMPTED



RATIO OF CORRECT TO ATTEMPTED

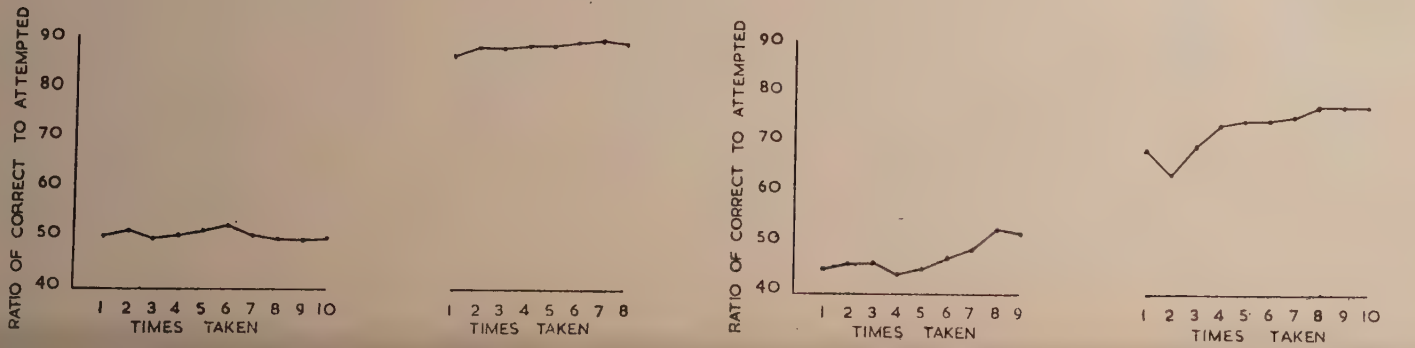


FIG. 3

unlimited time testing. In fact, it was found that the test was virtually an untimed test for the *duller* group after some 3 or 4 testings (see Figure 3), nearly all the Naval Ratings reaching the end of both Parts with several minutes to spare.

Both groups of subjects were asked to introspect about the test, in particular after their 3rd and last testings. They were encouraged to give any comments that occurred to them but were asked especially about the role of memory, awareness of boredom, preference for either Part over the other, and information as to how much they thought about the test and discussed it between testing sessions. Examination of these data did not yield any new or suggestive information: in the interests of brevity, therefore, we have omitted to describe or discuss them.

III

Results of Experiments with AH 5

The General Trend of Scores.

Table A shows the mean scores at successive testings of the 2 groups who took Test AH 5. Comparison of these scores with those produced by the earlier AH 4 subjects (Heim and Wallace, 1949, 1950), yields several points of interest.

TABLE A
AH 5 Repeated Retesting
Mean Scores

Times taken	*Technical College*			*Naval Ratings*		
	Part I	*Part II*	*Total*	*Part I*	*Part II*	*Total*
1	13·5	14·9	28·4	9·5	10·8	20·3
2	16·7	15·7	32·4	11·4	15·1	26·5
3	19·1	18·9	38·0	11·9	16·9	28·8
4	21·0	21·2	42·2	11·5	16·1	27·6
5	21·5	23·5	45·0	12·5	17·4	29·9
6	22·9	23·5	46·5	13·5	17·6	31·1
7	23·1	25·3	48·3	15·1	17·9	33·0
8	24·1	26·2	50·3	16·5	19·1	35·6
9	24·1	27·1	51·2	15·5	20·0	35·5
10	24·7	27·3	52·0			

All the AH 5 mean scores tend to show a steady, slow rise. Thus, of the 4 groups tested on AH 4 and AH 5, the only one showing a marked ceiling effect is the group of intelligent subjects (W.E.A.) on a test which is well within their scope. This effect is not seen with dull, with average or with intelligent subjects on tests which fully extend them. The AH 5 groups show neither the flattening towards the end of the experiment nor the startling rise towards the beginning. This suggests very strongly that any flattening of the progress curve found in repeated retesting is likely to be due to *relative* ease of the test, after some trials, rather than to some intrinsic property of "learning," in intelligence test performance.

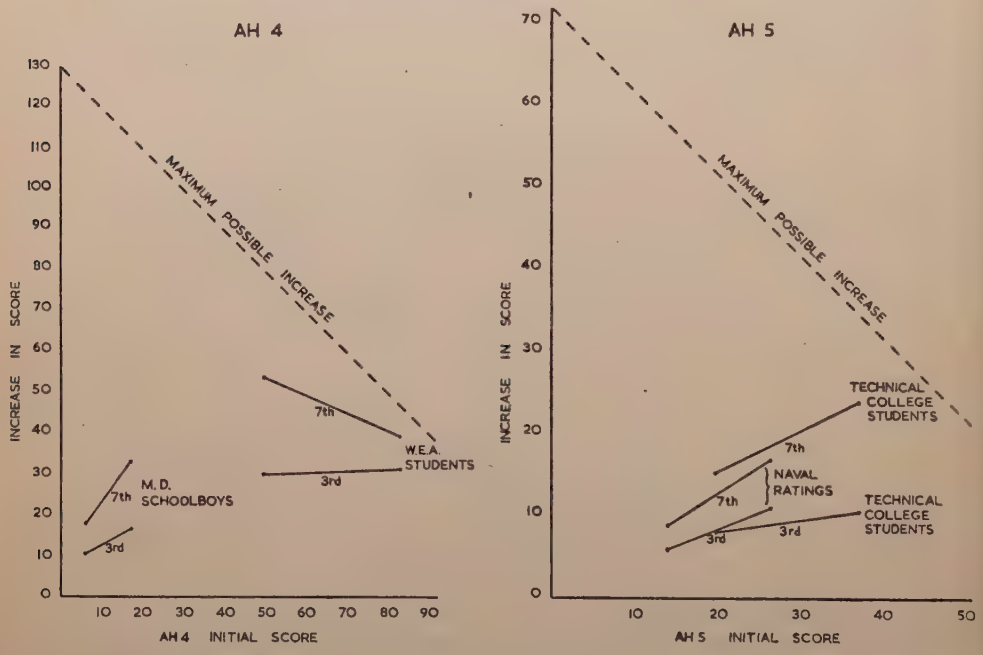
The progress curves of individuals, shown in Figures 1 and 2, bear out this hypothesis: none of the progress curves of the Naval Ratings show signs of flattening in the later trials but the top 3 or 4 subjects from the Technical College do show such

signs. In general trend, both sets of curves on the difficult test resemble those of the worst 8 of the mental defectives rather than those of the W.E.A. students on AH 4.

Inspection of the 4 sets of individual progress curves reveals that the mental defectives are by far the most erratic performers. They are the only subjects who show frequent, sudden, unpredictable drops followed by equally sudden rises. The extremely erratic progress of the individual mental defectives as compared with members of the other 3 groups suggests that smoothness of test performance over several trials is probably determined by absolute level of ability rather than by the relation between this and the test used. This is, of course, relevant to the question of consistency of subjects within their group, which is discussed in the following section.

FIG. 4

THE RELATION BETWEEN IMPROVEMENT AND INITIAL SCORE AT THE 3rd TESTING AND AT THE 7th TESTING



On both tests, it was the more intelligent subjects who yielded the highest standard deviations, although of course on AH 4 the range of the brighter group was forced to decrease over the testings, whilst that of the duller group increased. On AH 5, however, both groups had the opportunity, and did show a tendency to increase their range in the course of the experiments.

Consistency.

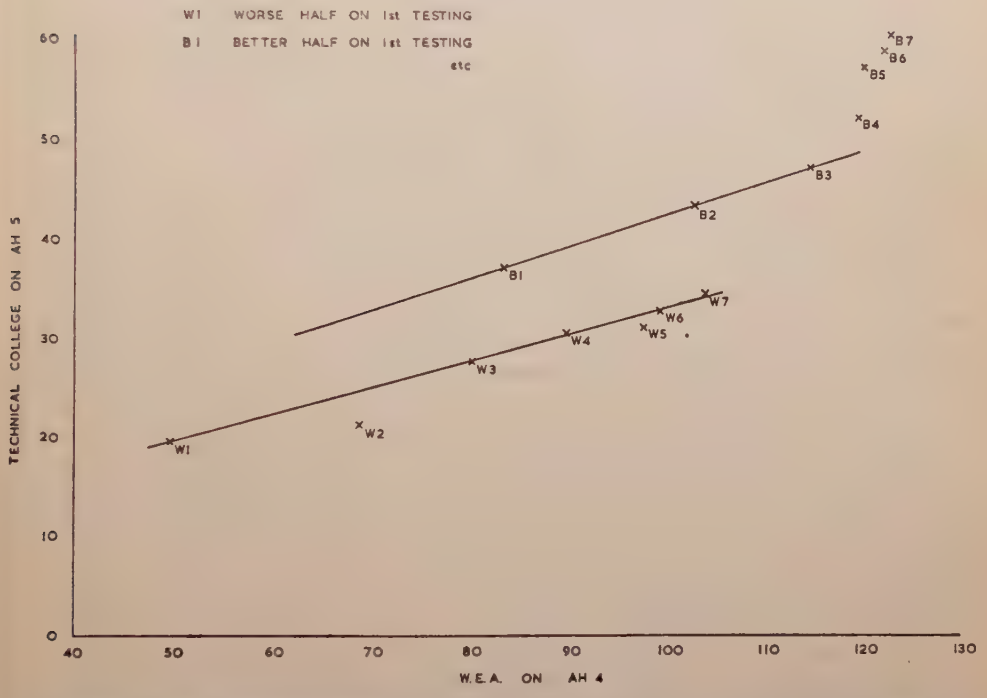
The product-moment correlation between every pair of testings was calculated for each of the 4 groups. Since these were uniformly high the matrices of inter-correlations are not given. With the exception of one pair of testings (the 2nd with

the 7th in the mentally defective group), r in every case exceeded 0·8 and, in most cases, exceeded 0·9. This suggests that the effects of individual variation, within the groups, are surprisingly small. However, this high "consistency" of the tests cannot be assumed in non-experimental situations, in which subjects are differentially test-sophisticated (see page 193).

Comparison of the 4 sets of matrices reveals 3 points of interest. First that the mentally defective subjects on AH 4 yielded the lowest sets of correlations. This is very likely due to the combination of their low intelligence (with the subsequent approach to random answering) and their temperamental lack of stability.

FIG. 5

COMPARISON OF AH4 AND AH5 BY EQUATING W.E.A. AND TECHNICAL COLLEGE GROUPS



Secondly, the highest values of r, in all 4 matrices tend to be along the diagonal. This is notably so with the mentally defective group. Thus there is a tendency for the results of adjacent testings to resemble one another more closely than do the results of more spaced testings.

Thirdly, the values of r approach 1·0 towards the end of the experiments, suggesting that the rank order of the subjects becomes still more stable after several testings. This agrees with the impression given by the individual progress curves.

None of the differences between correlation coefficients is significant, however.

Effects of Bias.

It may be remembered that the biases in Tests AH 4 and AH 5 are similar: both tests have a verbal and numerical bias in Part I and a visual bias in Part II.

It was thought interesting to compare the relationship between the 2 Parts for the 4 groups, in an attempt to ascertain whether bias exerts a differential effect in accordance with the relation between the test and the group.

The mean W.E.A. curves on the 2 Parts of AH 4 are almost exactly superimposed; those of the mental defectives are very widely separated, their Part II scores starting much higher and improving to a greater extent than their Part I scores.

AH 5 gives similar results; the difference between the progress curves of the 2 Parts is greater for the Naval Ratings than for the Technical College students, the visual Part proving the easier in both cases. In this respect, the 4 groups may be said to form a sort of series: intelligent subjects on AH 4 (no difference between Parts), intelligent subjects on AH 5 (slight difference), average subjects on AH 5 (greater difference), mental defectives on AH 4 (greatest difference). Thus the greater relative ease of the visually biassed questions increases as we descend the intelligence scale—or, rather, it again is a property of the relation between the test and the group tested.

The correlations between Parts I and II of the tests for the 4 groups were calculated. There was no definite trend except in the case of the W.E.A. group, for which the correlation dropped continuously from the 4th testing. This was thought, however, to be probably another result of the ceiling effect rather than a sign of the increasing effect of bias, and examination of the absolute difference between the scores on the 2 Parts confirmed this. In fact, the average absolute difference dropped with repeated retesting, so that the effect of bias if anything grew less. Moreover, although individuals in the other 3 groups showed a fairly consistent tendency to score higher in either Part I or Part II, in the W.E.A. testings there was a good deal of oscillation.

The correlation of Part I with Part II was far lower throughout for the mental defectives than for any other group. This could be partly but not entirely explained by their random answering on Part I, for correlations between successive scores on Part I were also fairly low, although those on Part II were high. Thus bias seems to exert a slightly greater effect on dull subjects, in the sense that the score on one Part of the test is less related to that on the other. Moreover, a visually biassed test seems to discriminate between such subjects much more than does a verbal and numerical test.

Questions Attempted and Answers Correct.

Figure 3 shows for each group the mean number of questions attempted throughout the testings, and the ratio of correct answers to questions attempted, expressed as a percentage. It may be seen that whereas the number attempted tended to increase throughout the testings, the ratio of correct to attempted remained virtually stationary for both the AH 4 groups—at a high level for the W.E.A. subjects and at a low level for the mental defectives. The W.E.A. students found the test too easy, too early to allow for much improvement, even the number attempted ceasing to increase much after the 5th testing. The mental defectives, on the other hand, apparently found the test too *difficult* to effect much improvement: whilst their number attempted (very low to start with) rose quite steadily they, like the White Queen, evidently had to exert all their energy to keep in the same place.

We have a very different picture when we turn to the AH 5 graphs. Here it is the *more* intelligent subjects whose number attempted started lower and at the end was still rising steadily. The steep initial rise and subsequent flattening of the Naval Ratings' graph suggest that their ambition was to reach the end of the test

at all costs: they did not realize the difficulty of the problems or they cared less about correctness than speed. A comparison of the 2 Naval Ratings' graphs confirms these hypotheses: it may be seen that the *ratio* begins to rise only when the graph of *attempted* flattens, i.e. the majority of the Naval Ratings having attempted nearly every question, they then settled down to answering these questions correctly.

A comparison of the Naval Ratings with the Technical College subjects shows clearly the greater speed of the former and the greater "power" of the latter (if "speed" be defined without any reference to correctness). The Naval Ratings always attempted more questions on the average than did the Technical College students. This, combined with the far higher ratio graph of the students confirms the hypothesis of the appropriateness of Test AH 5 for the latter. They evidently appreciated the difficulty of the problems, took their time and so managed to improve on their initially high ratio of correct to attempted.

The pair of Technical College graphs show just what would be expected of intelligent subjects repeatedly taking a test which fully extends them. Their ratio rises slowly and steadily (if we ignore the enterprising subjects who worked backwards on the 2nd trial) whilst their number attempted rises equally steadily but rather faster.

IV

Discussion

The chief reason for repeating the same type of experiment with groups and tests of different levels was in order to discover how far results which had hitherto been attributed solely to "practice" might in fact be dependent on the relation between the group and the test. Such effects as, for instance, that worse subjects improve more on retesting are suspect if the range of scores lies towards the top of the scale in the 1st testing.

When one particular question—that of relative improvement of better and worse subjects—was being considered it was found convenient to split each group into an upper and lower half on the criterion of initial score. It was found that the behaviour of these subgroups shed some light also on certain other problems, for instance, on whether some of the results in the 1st paper were due to the relative ease of AH 4 for the W.E.A. group. We shall begin, therefore, by considering some results for these subgroups.

TABLE B

All groups split in half on initial score

Test	*Group*	*Half*	*Average*			*Increase*	
			1st Score	*3rd Score*	*7th Score*	*3rd minus 1st*	*7th minus 1st*
AH 4	Mental Defectives	Worse	6·0	16·5	23·7	10·5	17·7
"	" "	Better	17·2	33·8	50·3	16·6	33·1
"	W.E.A.	Worse	49·7	79·7	103·8	30·0	54·1
"	"	Better	83·0	114·3	122·3	31·3	39·3
AH 5	Technical College..	Worse	19·7	27·7	34·9	8·1	15·2
"	" " ..	Better	37·3	47·9	61·1	10·6	23·8
"	Naval Ratings ..	Worse	14·0	20·0	22·8	6·0	8·8
"	" " ..	Better	26·5	37·5	43·3	11·0	16·8

Each of the 4 groups of subjects was arranged in order on initial score, and the top half was taken as one subgroup and the lower half as another. (When the total number was odd, the middle subject was omitted.) For each subgroup the average score on the 1st, 3rd and 7th trials was computed. The 7th was chosen rather than the last because the total number of testings varied with the groups: and the 3rd was chosen rather than the 2nd because of the anomalous behaviour of the Technical College students at the 2nd trial. The average increases between the 1st and 3rd and between the 1st and 7th testings were then calculated (see Table B). For each group of subjects it was found that the increase from 1st to 3rd testing was greater for the better subgroup than for the worse. Moreover, the same effect was found between 1st and 7th testing *with the exception of the W.E.A. group.*

It is, of course, clear that equal increments in score on different parts of the scale do not necessarily represent equal increments of improvement; but at least the consideration of the complete subgroup will tend to average out local affects of variability of scale. Do the results in Table B merely reflect some major discontinuity of the scale? We think not. If we set aside the W.E.A. for a moment, we have 3 groups either on different tests or centred on different parts of the scale of the same test, and each showing the same tendency.

The results of Table B are shown in Figure 4. The line marked "M.D. schoolboys 7th," for instance, is obtained as follows: the average increase in score between 1st and 7th testing is plotted against average initial score for the worse half of the mental defectives and also for the better half of the same group, and these 2 points are joined. The line obtained gives an estimate of the way in which initial score is related to improvement in score. If the line were horizontal it would show that the improvement was the same whatever the initial score; if the line sloped upwards, the initially brighter subjects, and if it sloped downwards, the initially duller subjects, would show most improvement. In fact (except for the W.E.A.), not only are all these lines sloping upwards, but the slope for the 7th testing is greater than for the 3rd, so that it appears that in general the better people improve more than the worse to begin with and continue to improve more rapidly with successive testings: they "learn" more by practice. With the W.E.A. students there is a striking reversal of this effect, the 1st to 3rd line being almost horizontal and the 1st to 7th even sloping downwards.

The cause of this discrepancy is demonstrated by the rest of the Figure. The dotted line indicates the greatest increase possible at any given score (for instance, if the original score is 60, the maximum possible increase is 70, since the maximum score on the test is 130) and there can be no doubt that it is this which deflects the line "W.E.A. 7th" and even has some effect in depressing "W.E.A. 3rd."

We may therefore conclude that the greater improvement of the worse half of the W.E.A. group is caused by the fact that many of the W.E.A. are too near the top of the test. The effect of practice on this group has been to carry some of them outside the range of the test altogether.

In general then we may say that the better subjects improve with practice more than the worse subjects except where the scope for increase for the better subjects is limited. It is necessary to emphasize here that we have used as the criterion of "better" and "worse" the subject's initial score in the test. This may or may not be a measure of the subject's intelligence—in the case of a test which is too easy for a subject it will be largely a measure of speed—but it seems as though the effect will occur whether it is or not. However, it is when the test is both too easy and

too short that the ceiling effect will operate, for then the highest initial scorers will have least chance of improving.

The results of our 4 experiments indicate that the effect of test sophistication may be considerable and may be misleading if the subjects are not equally practised. Let us take as illustration the most striking case possible in which the worse half of each set has between 7 and 10 practice trials and the better half has none. For the worse subgroup, then, we take final scores and for the better subgroup we take initial scores, obtaining the following results for the 4 groups.

	W.E.A.		*Mental Defectives*		*Naval Ratings*		*Technical College*	
	Apparently		*Apparently*		*Apparently*		*Apparently*	
Initially	*Worse half*	*Better half*	*Worse half*	*Better half*	*Worse half*	*Better half*	*Worse half*	*Better half*
Worse half	1	5	1	5	2	2	$2\frac{1}{2}$	$4\frac{1}{2}$
Better half	5	1	5	1	2	2	$4\frac{1}{2}$	$2\frac{1}{2}$

Thus, in a selection procedure we should tend to get the "wrong" subjects above the median except in the case of the Naval Ratings, when selection would be random.

These results are based on repeated testing with the same test, at frequent intervals over a short period and it is, of course, by choice the most far-fetched case. If we find similar results when using parallel, but not identical, tests, it would suggest that considerably more precautions should be taken in the practical application and interpretation of test scores.

The form which such precautions should take is not easy to determine: in selection for secondary schools, for instance, if test results are used at their face value, either as a means of initial screening or as one of the final criteria, the individual child who has escaped test practice will clearly be penalized. On the other hand, adoption of a scheme such as Valentine's (1948)—whereby the number of high-scoring children from each school determines the proportions to be accepted from each school—would tend to favour the school most energetic in practising tests. These difficulties may explain but they hardly excuse those educationalists who have ignored, or minimized, the practical importance of test sophistication in mental testing.

When considering the effects of bias and the relative ease of Part II for the 2 duller groups, it occurred to us that this effect might be due to all members of these 2 groups being male, since the tendency has been found always for male subjects to score relatively highly on non-verbal material. However, comparison of the scores gained by men and women on the 2 other groups did not confirm this hypothesis: among the more intelligent subjects, no "preference" was found for one Part over the other.

These findings are confirmed by the results of an item analysis of AH 5 (now in progress) which suggests that the greater ability of boys and men to deal with "visual" material may be confined to low-scoring subjects. It would appear then that the discrepancy of score between the 2 Parts shown by the mental defectives and the Naval Ratings is due to their being both male and relatively unintelligent.

The Graphs in Figure 3 showing the number of questions attempted and the ratio of correct answers for each group, throw some light on the problem of what may be meant by "power," "speed" and "accuracy" in mental testing. Perhaps

we should say rather that they add depth to the already dark shadows veiling this topic, for comparison of the graphs one with another makes it clear that in so far as these terms have any significance, it is purely relative. The same test may be for one group a test of "speed" and for another, a test of "power"; according to its position in a series of testings, the same test for the same group may be mainly a test of "accuracy" or mainly a test of "speed"; and the distinction between "power" (as opposed to "speed") and "accuracy" (as opposed to "carelessness") is clearly untenable when the same subjects, failing certain questions in trials 3 or 4, answer them correctly in trials 7 or 8.

We are not suggesting that these terms are valueless or that they should be excluded from the mental tester's vocabulary (to be replaced, in all probability, by other and longer terms). But we should question the value of assuming that such terms refer to absolute and unitary qualities, and that certain existing tests do measure these qualities.

In conclusion, we should like to indicate that we are aware of the criticisms which may be levelled at the use of mental test scores as a scale of measurement and realize that the results we have obtained are a function of the scale. However, there are no grounds for supposing that our tests yield scores of a fundamentally different kind from those of other intelligence tests: given the usual assumptions made by mental testers, we think our results are valid.

V

Conclusions

The most important results yielded by the 4 experiments seem to us to be as follows: that test scores in all circumstances tend to increase as the subjects become more practised; that this improvement is to some extent differential, since the initially high scorers, given scope, tend to show the greatest increase; and that, following from these results, the practical value of intelligence tests is liable to diminish when the members of the group tested differ in their degree of familiarity with such tests. The order of subjects within their group was found to be fairly stable throughout testings, this stability tending to increase towards the end of the experiments.

Visually biassed questions were found to be easier than verbally and numerically biassed questions, *for the less intelligent subjects*. No such uniform preference was manifested (on the criterion of initial score or of rate of improvement) by the more intelligent subjects.

Comparison of the number of questions attempted and the ratio of number correct to number attempted, for each group on its test, suggested some criticisms of the concepts of "power," "speed" and "accuracy" in mental testing.

These results suggest that:

(*a*) The improvement in score effected by test-sophistication has insufficiently attracted the attention of mental testers engaged in practical problems of selection, despite previous investigators in general having agreed on the fact of improvement. This may be because the consistency of the tests remains high over repeated retesting but this does not affect the case of the *differentially* sophisticated group.

(*b*) The conflicting results obtained by previous investigators are probably explained by the fact that many tests do not fully extend all members of a group after several testings and that investigators have tended to take a "static" rather than an "interactive" view of their tests and of their subjects.

(*c*) In view of the greater improvement shown by the duller subjects on the visually biassed test questions, it is desirable to use either tests which have several biases or batteries of tests each with a different bias. This is especially important in practical work with subjects of low intelligence who may be differentially test-sophisticated.

The writers wish to express their thanks to all subjects who took part in the experiments; to the staff of the Cambridgeshire Technical College for their help in attracting subjects and in providing accommodation for testing; and to Miss V. Horn for drawing the figures.

References

1. Dave, K. (1938). Effects of practice upon intelligence tests. *Brit. J. educ. Psychol.*, **8**, 313.
2. Dawson, S. (1936). The influence of practice on the results of mental tests. Eighth Annual Report of the Scottish Council for Research in Education.
3. Dunlap, K., and Snyder, A. (1920). Practice effects in intelligence tests. *J. exp. Psychol.*, **3**, 396–403.
4. Harris, N. (1947–9). Reports on testing and retesting grammar school girls at yearly intervals. Unpublished.
5. Heim, A. W. (1947). An attempt to test high-grade intelligence. *Brit. J. Psychol.*, **37**, 2, 70–81.
6. Heim, A. W., and Wallace, J. G. (1949). The effects of repeatedly retesting the same group on the same intelligence test: I, Normal adults. *Quart. J. exp. Psychol.*, **1**, 4, 151–159.
7. Heim, A. W., and Wallace, J. G. (1950). The effects of repeatedly retesting the same group on the same intelligence test: II, High-grade mental defectives. *Quart. J. exp. Psychol.*, **2**, 1, 19–32.
8. McIntosh, D. M. (1944). The effect of practice in intelligence test results. *Brit. J. educ. Psychol.*, **14**, 44–45.
9. McRae, H. (1942). The inconstancy of group test I.Q.'s. *Brit. J. educ. Psychol.*, **12**, 59–70.
10. Richardson, F., and Robinson, E. S. (1921). Effects of practice upon the scores and predictive value of the alpha intelligence examination. *J. exp. Psychol.*, **4**, 300–317.
11. Rodger, Allan G. (1936). The application of six group intelligence tests to the same children, and the effects of practice. *Brit. J. educ. Psychol.*, **6**, 291–305.
12. Thorndike, E. L. (1922). Practice effects in intelligence tests. *J. exp. Psychol.*, **5**, 101–107.
13. Valentine, C. W. (1948). Some present-day trends, dangers, and possibilities in the field of psychology. *Brit. J. educ. Psychol.*, **18**, 3, 143.
14. Vernon, P. E. (1938). Intelligence test sophistication. *Brit. J. educ. Psychol.*, **8**, 237–244.
15. Wallace, J. G. (1949). Results of a high-grade intelligence test applied to a university population. Unpublished.
16. Woodrow, H. (1938). The effects of practice on groups of different initial ability. *J. educ. Psychol.*, **29**, 4, 268–278.

(*Received 20th May*, 1950.)

STATISTICAL APPENDIX

I

Structural Relation

The method of estimating the relation between the initial score and the increase in score by splitting each group into a worse and a better half is that advocated by A. Wald (*Annals of Mathematical Statistics, II*, 1940, p. 284). He points out that, where both the dependent and the independent variable are subject to error, a regression line does not

give the best estimate of structural relation, although it is of course still the best method of predicting values of the dependent variable from given values of the independent variable. For instance, suppose there is a true relationship between the variables of the form $y = ax + b$, which in this case would mean that if we know the initial score (x) we could determine exactly the true increase in score (y) at, say, the 3rd testing; and suppose that x and y are distorted by random errors so that for the i^{th} subject we have score $x_i = x + \epsilon_i$

and increase $y_i = y + \eta_i \qquad i = 1, 2, \ldots . 2n$

instead of x and y. Then as Wald shows, a can be estimated by

$$\frac{\frac{y_1 + \ldots . + y_n}{n} - \frac{y_{n+1} + \ldots . + y_{2n}}{n}}{\frac{x_1 + \ldots . + x_n}{n} - \frac{x_{n+1} + \ldots . + x_{2n}}{n}}$$

where the first n subjects are those with the lowest x's; and this will not be the same as the slope of the regression line of y_1 on x_1. This estimate has the merit that it gives the same line whether we are enquiring into the relationship of x to y or the relationship of y to x, whereas the two regression lines are different except when the correlation coefficient $r = 1$, a fact which can lead to much confusion in this sort of work.

It is clear that in this case it is the structural relation that we require, not an equation for prediction. We are asking the general question: do better subjects improve more or less than worse ones?—not the particular one: given an individual subject's initial score, what is his score at a later testing likely to be?—and so Wald's method is the appropriate one.

The underlying assumption of this method, that the error variance is the same at all points of the scale, cannot be tested and is unlikely to be true for intelligence test scores. We have had to rely on the argument that we get similar results for different groups either in different tests or centred about different parts of the scale of the same test, to show that the estimate is roughly correct. We have also used the similarity of the results as an excuse for omitting the calculation required to decide whether the worse half of the group had been accurately selected by taking the lower initial scores, which are subject to error.

The slopes of the lines shown in Figure 4 were tested to see if they were significantly different from 0 (the significance test is equivalent to a t-test on the average increase in score of the 2 subgroups). None of the lines for the 3rd testing had a significant slope, but all those for the 7th testing had, except that of the Naval Ratings. The subgroups are, of course, very small.

II

Comparison of AH 4 and AH 5 by Equating W.E.A. and Technical College Groups

We have shown that the relation between the W.E.A. group and its test is different from that between the other groups and their tests, and have given reasons for thinking that this is because AH 4 is too easy for them. It might perhaps have been desirable to give the same group repeated retesting on both tests, but, apart from the difficulty of keeping a group together for long enough to finish such an experiment, there would have arisen certain problems with regard to the transfer of training which would have made the interpretation of results far from straightforward, and there is much to be said for the method we have used for demonstrating the effect of the relation of group to test. What we have lost however is the possibility of direct comparison between the rates of progress on the two tests.

In an attempt to remedy this loss, we decided to equate the Technical College and W.E.A. groups, which are in fact similar in occupation, interests and general background and therefore probably similar in level of intelligence. We graphed the average score on AH 4 of the worse half of one group against the average score on AH 5 of the worse half of the other (and similarly for the two better halves) for each of the first 7 testings (see Figure. 5). It will be seen that the 7 W plots (of worse half-groups) lie very close to a straight line, and that the first 3 B plots lie on another which is parallel to, but not identical with, the first.

The position of B 4 to B 7 shows that the rate of improvement slows down on AH 4 before it does on AH 5—the effect of the end of the test, as usual—while the fact that W 1 to W 7 are almost collinear shows that, where the end of the test does not interfere, the relative rate of improvement on the two tests remains the same. The latter result is rather unexpected as there might well have been a stage in AH 5 at which it became much harder for the worse subjects, in which case the W points would have dropped below the line. It seems therefore that for groups of this level, the gradient of difficulty for AH 5 is not very different from that for AH 4.

It might be thought that the B and W lines should coincide; it is worth considering why they do not. If they had done so it would have shown that a given score on AH 4 could be equated with a score on AH 5, and vice versa, whether this score were obtained at a 1st testing or at a later one. Since the B line is above the W line, either this is not true, or there is a greater gap between the better and worse halves of the Technical College group than between those of the W.E.A. However, the ratio of the mean difference of the subgroups to the standard deviation is 3·0 for the Technical College and 3·8 for the W.E.A. so that the gap is if anything smaller for the Technical College, and we must conclude that, if we try to convert a score on one test into a score on the other the number of previous testings, as well as the actual score, must be taken into account, i.e. the effect of practice on the two tests is different. In fact, since W 3 is almost below B 1, by the 3rd testing the worse half have nearly equalled the better half on AH 4, but since it is W 7 that is in horizontal line with B 1, it is not until the 7th testing that the worse half reaches approximately the level of the better half on AH 5.

BOOK REVIEW

Human Ability. By C. SPEARMAN, Ph.D. (Leipzig), Hon.LL.D. (Wittenberg), F.R.S., and LL. WYNN JONES, M.A. (Oxon.), Ph.D. (Leipzig), D.Sc. (Wales). London: Macmillan & Co., Ltd., 1950. Pp. vi + 198. 16s.

This book is both interesting and disappointing. Its interest is largely historical since, as Wynn Jones tells us in his Preface, Spearman planned it to be a continuation of his "Abilities of Man"; moreover, the book refers extensively to writings whose appeal is to the historically-minded only. Its disappointing quality is most striking in the light of the authors' Concluding Remarks. Here we read, with approval and hope, that "no one can say that a final interpretation of *any* broad factor has been obtained" and that "an ultimate explanation . . . is still anybody's guess." But this air of caution and fluidity is not justified by the body of the book.

"Human Ability" consists largely of an inventory of the work of many psychologists whose views were different from Spearman's; but no recognition is accorded the underlying presuppositions (which they share with Spearman), and no space is devoted to those whose views were so fundamentally *other* that they share virtually no common ground with the authors of "Human Ability." For instance, Binet (apart from two brief mentions in the first chapter, entitled "Pre-Scientific Notions") is dismissed in the body of the book with one comment: "It is well known that the various speculations of Binet as to the nature of intelligence were hardly consistent"! Few writers on psycho-pathology would be content to limit their evaluation of Freud to such a reflection.

The problems treated concern for the most part such questions as the number of acceptable factors, the extent of their interdependence, the eventual compatibility of the various statistical methods which have yielded incompatible results; and emphasis is persistently laid on the exactitude of the computations, the impressive numbers of subjects participating and the variety of tests which have been used. The following excerpts probably give an idea of the writers' sense of values more briefly and clearly than could be achieved by the reviewer: "Perhaps the most outstanding advance in this direction was achieved by Brown and Stephenson who in 1932 not only made use of twenty different tests but, furthermore, submitted the results to mathematical treatment of previously unparalleled rigour, fundamentality, and laboriousness. Almost simultaneously, however, even all this was far out-topped by the aforesaid Chicago work, wherein the number of tests rose to no less than ninety-nine whilst the mathematical treatment was so exhaustive as to be still in progress." "Even here, however, the evidence has gone to show that such a factor was not so much really absent as only latent. By means of appropriate statistical procedure it could be immediately revived in full force." "In mental as in physical science, however, the characters most satisfactory for fundamental developments must . . . be replaced by others that are more tractable. Thus the physical measurements purporting to measure 'energy,' 'ether,' and so forth, actually derive their content from visual percepts such as those of thermometers or galvanometers. In similar wise, the mental measurements are shy of grappling with the illusive 'clearness' and have instead recourse to more or less indirect indications."

The writers make convincing criticisms of Thurstone, notably (in Chapter X) on the question of testing "the power of abstraction," but they appear unaware that such criticisms could be applied with equal cogency to some of their own conclusions—in particular those which depend on taking the name of a test at its face value. Perhaps the most startling jump from a name to a "fact" is to be found in the chapter on Other Broad Qualitative Factors: ". . . the noegenetic theory has taken a hand, by showing that the essential form of all mental creation or invention consists in educing correlates."

The matter of naming tests brings us to one of the fundamental problems of the factor analyst—or rather of his critics. "What's in the name?" they ask, "What is meant by 'G' and 'p' and 'O'?" "Nothing," would appear to be the answer if the critics look for instance at page 25: "Not for us . . . is the luxury of exploring the depths of ontology and epistemology. We are under no obligation to answer such questions as: Whether 'factors' have any 'real existence'? do they admit of genuine 'measurement'? does the notion of 'ability' involve at bottom any kind of cause, or power?" (It seems odd that the last question is not for the writers, in view of the title of the book.)

This viewpoint, though sterile, would be unexceptionable if it were maintained consistently. Elsewhere, however, very definite implications are made as to the existence, the measurability and the meaning of factors. On page 111, for example: "Failing such an assumption . . . they would present something not susceptible of genuine measurement. . . . In order to extricate the tests from all this embarrassment . . . the sole effective resource is to pass from the ancient concept of faculties to the very modern one of 'factors.' . . . The first [step] is to determine a statistical coefficient or 'loading.' The second step is to supply this determination with an interpretation."

In this excerpt may be found the source of many of the doubts and fears which give pause to the unconverted. These suspect that the few factors which are generally accepted by all mental testers are precisely those (under new names) postulated by the "ancient" faculty psychologists; that the "very modern" factor analysts have lost in psychological understanding what they think to have gained in mathematical precision; and that the apparent modesty and objectivity of naming factors symbolically—relegating "interpretation" to a separate chapter—cloaks an inconsistency at least as heinous as that of Binet.

A. W. H.

ERRATUM.—In: Berlyne, D. E. Stimulus intensity and attention in relation to learning theory, the results of Experiment 1 should read (*page* 73):

Mean responses to larger stimulus: 12·67
" " " smaller " : 3·33
Number of subjects: 6.
$\sigma^2 = 3{\cdot}16 \quad t = 5{\cdot}88 \quad \nu = 5 \quad p = 0{\cdot}002.$

INDEX OF AUTHORS

Page

BARTLETT, F. C.
Programme for experiments on thinking - - - - - - - 145

BELBIN, E.
The influence of interpolated recall upon recognition - - - - 163

BERLYNE, D. E.
Stimulus intensity and attention in relation to learning theory- - - 71
Erratum - - - - - - - - - - - - - 199

CANE, V. R., with Heim, A. W.
The effects of repeated retesting: III. Further experiments and general conclusions, *with statistical appendix* - - - - - - - 182

CARPENTER, A.
A comparison of the influence of handle load and of unfavourable atmospheric conditions on a tracking task - - - - - - - 1
A simple class-room chronoscope - - - - - - - - - 140

DAVIS, D. RUSSELL, with Sinha, D.
The effect of one experience upon the recall of another - - - - 43
The influence of an interpolated experience upon recognition - - - 132

DEUTSCH, J. A.
A new phenomenon of apparent movement and its after-effect - - 119

FRASER, D. C.
The relation between angle of display and performance in a prolonged visual task - - - - - - - - - - - - - 177

HALL, K. R. L., with Oldfield, R. C.
An experimental study on the fitness of signs to words - - - - 60

Page

HALL, K. R. L.
Perceiving and naming a series of figures - - - - - - - 153

HEIM, A. W., with Wallace, J. G.
The effects of repeatedly retesting the same group on the same intelligence test: II. High-grade mental defectives - - - - - - - 19

HEIM, A. W., with Cane, V. R.
The effects of repeated retesting: III. Further experiments and general conclusions - - - - - - - - - - - - - 182

HICK, W. E.
The threshold for sudden changes in the velocity of a seen object - - 33

HOPKINSON, R. G.
The multiple criterion technique of subjective appraisal - - - - 124

OLDFIELD, R. C., with Hall, K. R. L.
An experimental study on the fitness of signs to words - - - - 60

POULTON, E. C.
Perceptual anticipation and reaction time - - - - - - - 99

SINHA, D., with Davis, D. Russell.
The effect of one experience upon the recall of another - - - - 43
The influence of an interpolated experience upon recognition - - - 132

SZAFRAN, J., with Welford, A. T.
On the relation between transfer and difficulty of initial task - - - 88

TSAO, J. C.
Studies in spaced and massed learning: III. Paired associate and serial learning - - - - - - - - - - - - - 13

WALLACE, J. G., with Heim, A. W.
The effects of repeatedly retesting the same group on the same intelligence test: II. High-grade mental defectives - - - - - - - 19

WEISKRANTZ, L.
Figural after-effects in stroboscopic motion - - - - - - - 113
An unusual case of after-imagery following fixation of an "imaginary" visual pattern - - - - - - - - - - - 170

WELFORD, A. T., with Szafran, J.
On the relation between transfer and difficulty of initial task - - - 88

WHITFIELD, J. W.
The imaginary questionnaire - - - - - - - - - 76
Note on the correlation of errors and performance in a test - - - 138

WILLMER, E. N.
The relationship between area and threshold in the central fovea for lights of different colours - - - - - - - - - 53

ZANGWILL, O. L.
Amnesia and the generic image - - - - - - - - - 7